



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Forecasts: Fact or Fiction?

Uncertainty and Inaccuracy in Transport Project Evaluation

Nicolaisen, Morten Skou

Publication date:
2012

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Nicolaisen, M. S. (2012). *Forecasts: Fact or Fiction? Uncertainty and Inaccuracy in Transport Project Evaluation*. Department of Development and Planning, Aalborg University.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Forecasts: Fact or Fiction?

Uncertainty and Inaccuracy in
Transport Project Evaluation

Ph.D. Thesis by Morten Skou Nicolaisen

Department of Development and Planning - Aalborg University

December 2012

Til mine forældre

Forecasts: Fact or Fiction? Uncertainty and Inaccuracy in Transport Project Evaluation

Ph.D. Thesis

2. Edition (online version)

December 2012

© Aalborg University and Morten Skou Nicolaisen

Department of Development and Planning, Aalborg University

Author: Morten Skou Nicolaisen

Supervisor: Petter Næss

Co-supervisor: Arvid Strand

Cover: Silvia Dragomir

Printed by: Uniprint, DK-9210 Aalborg

Total pages: 226

Contents

PREFACE	1
DANSK RESUMÉ	4
ABSTRACT	5
1 INTRODUCTION	6
1.1 TRANSPORT PLANNING AND THE NEED FOR APPRAISAL	6
1.2 COST-BENEFIT ANALYSES LEAD THE WAY	9
1.3 PROBLEM STATEMENT AND RESEARCH QUESTIONS	11
1.4 SUMMARY	13
2 BASIC CONCEPTS	15
2.1 CAUSALITY IN A STRATIFIED REALITY	15
2.2 SOCIAL SCIENCE AND MODELS	20
2.3 CLASSIFYING UNCERTAINTY IN MODEL-BASED DECISION SUPPORT	25
2.4 ACCURACY AND THE ROLE OF PLANNING	28
2.5 FOUR STAGE MODELS	31
2.6 SUMMARY	35
3 RELATED WORKS	36
3.1 POST-AUDITING OF APPRAISAL	36
3.2 STUDIES OF FORECASTING ACCURACY	40
3.3 SYNTHESIS OF FINDINGS IN PRIOR ART	61
3.4 SUMMARY	64
4 METHODOLOGY	66
4.1 ADDRESSING THE PROBLEM STATEMENT	66
4.2 DEGREES OF INACCURACY IN FORECASTING	66
4.3 PLANNING IN THE FACE OF UNCERTAINTY	74
4.4 DATA SOURCES	78
4.5 SUMMARY	80
5 DATA COLLECTION	81
5.1 DECISION SUPPORT DOCUMENTS	81
5.2 DATABASES	84
5.3 QUESTIONNAIRES	88
5.4 INTERVIEWS	90
5.5 SUMMARY	91
6 FORECASTING INACCURACY	92
6.1 FORECASTING ACCURACY	92
6.2 EXPLORING COMMON PROJECT TRAITS	106
6.3 THE SENSITIVITY OF IMPACT ASSESSMENTS	124
6.4 SUMMARY	136

7	UNCERTAINTY MANAGEMENT	137
7.1	UNCERTAINTY IN DECISION SUPPORT DOCUMENTS	137
7.2	SURVEY RESULTS ON UNCERTAINTY MANAGEMENT	143
7.3	INTERVIEWS	166
7.4	SUMMARY	201
8	CONCLUSION	202
8.1	FINDINGS	202
8.2	DISCUSSION	207
9	BIBLIOGRAPHY	216

Preface

The work presented in this thesis is not only the result of the three years of laborious studies that is part of doing a Ph.D. The motivation for engaging with the quality of decision support in policy-making came already at the start of my master studies at the Urban Planning and Management programme at Aalborg University. This was in the fall of 2007, where I wrote a report on how to improve the quality of urban spaces together with fellow students Anne Bach and Nina Vogel. During this semester I found myself becoming increasingly concerned with how the quality of both public and private spaces was influenced by transport activities, and how a pressure to accommodate demands for increasing levels of traffic often had an adverse effect on this quality in urban environments. I therefore decided to turn my focus towards transport planning as my field of specialisation.

As luck would have it, the Danish Infrastructure Commission was finishing its report on recommendations for the future development of Danish transport infrastructure planning at the same time we were finishing our report on the quality of public spaces. The commission's report therefore provided a good point of departure for investigating the underlying rationales for Danish transport planning for our next project, and we decided to analyse the underlying assumptions in this report. With the addition of Jeppe Andersen to the group we were now four people, where I was in charge of critically scrutinizing the assumptions in the report as well as the various sub-reports that the main document referred to. While I had expected the commission to be politically biased, I also expected the technical analyses to be somewhat sober. However, the investigation quickly showed that indicators of political bias was present in these 'independent' analyses, and the assumptions in the report all indicated that a decision for how to approach Danish transport planning in the future had already been established. The task left for the commission was seemingly to construct an analysis that could justify this approach. The traffic analyses completely ignored that congestion issues can (and probably should) be approached in other ways than constructing more and wider motorways, and the result was a commission that ended up recommending a doubling in length of the Danish motorway system to counter future congestion problems. Transport related environmental concerns were handled by recommending a fast introduction of electric cars.

Shocked at the low quality of the commission's analyses and the lack of critical reflection in the 312 page document that made up the main report, I chose to do a set of scenarios for how to achieve sustainable mobility during my one year master thesis. The findings clearly showed that a transport planning approach as outlined in the commission's report would never allow Denmark to reach the ambitious environmental commitments we had been hailed for internationally during the past decade. However, as with the public spaces I had studied on my first master semester, the demand for increasing levels of traffic overruled most other concerns, and the results from the commission's analysis quickly became the de facto standard for the future development of Danish transport infrastructure. While I was disappointed at these findings, I also found myself fascinated with how a mass of assumptions, some based on overwhelming evidence and others no more than mere guesswork, were condensed into grand models to (supposedly) aid decision makers in selecting the

most appropriate course of action. No matter how poor the assumptions in the model were, it seemed the mere act of using a model was perceived as providing valid arguments, and it therefore became difficult for opponents to the suggested course of action to argue against it without providing model-based traffic analyses of their own.

I was thus pleased to be granted an opportunity to pursue these issues further through a Ph.D. fellowship that aimed to investigate systematic biases in travel demand forecasts. This brings us to the present thesis, which is part of a larger research project aimed at examining uncertainties in transport project evaluation (UNITE) funded by the Danish Council for Strategic Research. The project includes five work packages in total, of which my work has been part of the first of these, so that the data collected here can be used in the remaining work packages. I am confident that the information presented over the next couple of hundred pages will be of use not only in these additional work packages, but also for researchers, planners, and policy makers involved with transport planning in general. I have therefore tried to keep the technical jargon to a minimum and instead present the work in a readable form that still captures all the necessary details of scholarly inquiry.

The work presented herein builds on two separate parts of analysis; one that is mainly quantitative and another that is more qualitative. The quantitative part has required the construction of a sizeable database of relevant information about completed transport infrastructure projects. While I was still in the initial stage of my Ph.D., Bent Flyvbjerg at University of Oxford encouraged me to start gathering data for this part immediately, since he expected data availability to be a major obstacle for my project. With Bent being one of the most prominent planning researchers to previously conduct large scale investigations of forecasting inaccuracy, I figured he would have good reasons to be concerned about this. I was therefore thankful for the input, but at the time I didn't appreciate just how important this advice actually was. However, as I started collecting data immediately after this conversation, it quickly became clear that Bent had been quite accurate in his prediction. Reliable data was incredibly difficult to come by, and I am therefore grateful to the large number of people who made it possible to obtain the necessary documentation. I believe the resulting database to be the most extensive of its kind to date, with a methodology that has both improved weaknesses found in previous studies as well as allowed observation of interesting new patterns of forecasting inaccuracy.

For their aid in data collection I wish to thank (in no particular order): Niels Erik Moltved, Katrine Handberg, Henrik Nejst Jensen, Ulrik Winther Blindom, and Rasmus Larsen at Vejdirektoratet. Svend Aage Bonde and Michael Preben Hansen at Banedanmark. Jens Brix and Nadia Zøllner Jensen at Trafikstyrelsen. Peter Krabbe Skousen at Metroselskabet. Jesper Mølgaard at DSB. Tine Lund Jensen, Karsten Kirk Larsen, and Liv Mølgaard Mathiasen at Transportministeriet. Henrik Paag at Tetraplan. Louise Wootton and Poul Beecher at the UK Highways Agency. Sylvia Yngström Wänn, Andreas Fernholm, Pär Ström, Lennart Lennefors, Stephen McLearnon, Håkan Persson, and Anneli Juujärvi at Trafikverket. Ulf Lindblom and Anders Hedlund at the Swedish EIA Centre. James Odeck, Kjell Johansen, Niels Myhren, Bjørn Martin Alsaker, Arild Hegreberg, Harald Granrud, Magne Lerfaldet, Gunnar Stiberg, Karl-Erik Willadsen, Åsmond Hansen, Thorbjørn Hansen, and Yngve Wikeland at Statens Vegvesen. Anne Cathrine Bakke at Ruter. Martin Lund-Iversen at NIBR. Håkon Rasmussen at Bergen Kommune. Camilla Lundberg Berntzen at Skyss. I am sure I managed to forget some

additional people that offered me their help along the way, and for this I apologize. It is lack of memory, not of appreciation, that is the sole cause of this. For the qualitative part, I am indebted to the many people who took their time to answer the questionnaire as well as those who agreed to participate in lengthy interviews. For reasons of confidentiality I cannot disclose the names here, but I am grateful to everyone that helped provide important pieces of information for the work presented in this thesis.

In addition to the external people who have helped with data collection, I am also grateful to those of my colleagues who have provided interesting discussions, reflections, and comments along the way, and I could list most of the Department of Development and Planning at Aalborg University here if space allowed. I am especially appreciative for the help in collecting and managing all the information provided by Patrick Driscoll and Teresa Næss at Aalborg University. I have also been grateful for the good collaboration in the UNITE project, where Otto Anker Nielsen, Steen Leleur, Bo Friis Nielsen, Kim Bang Salling, Stefano Manzo, and David Meisch at the Danish Technical University have all been in ample supply of useful feedback and references throughout the project, and I can only hope the appreciation is mutual.

A special thank you goes out to Arvid Strand at the Institute of Transport Economics, who was my co-supervisor during the Ph.D. and have offered plenty of advice on practical, theoretical, and philosophical issues in the later stages of the process. This gratitude extends to his wife Vibeke Nanseth, as they were hospitable enough to be my hosts during the 3 months I lived and worked in Oslo. They were also kind enough to provide a delightful retreat in Ortigia, where I could finish my thesis in tranquil surroundings while Anne enjoyed the charms of Sicily.

An entire chapter could be dedicated to Petter Næss at Aalborg University, who has gone far beyond his duties as a supervisor on both my master and Ph.D. theses. Since I cannot hope to ever express my gratitude to him in words, I shall simply state that I am forever indebted to him for the many hours we have shared in both professional and social settings. I could not have wished for a better supervisor, mentor, colleague, or friend, and much of the quality in the work presented in this thesis comes from his dedication as a supervisor. That being said, the responsibility for any weaknesses is of course solely mine.

Last, but certainly not least, my close friends and family deserve enormous praise for supporting me when things were looking difficult. Especially Anne, who have experienced all the ups and downs along the way, and managed to push me back on track whenever I was feeling frustrated. As a transport planner working outside academia, her reflections on my work have been a continuous source of inspiration as well as providing a necessary reality check from time to time. She continues to amaze me in her ability to approach any problem with a positive attitude, and I consider myself extremely lucky to have her by my side.

Dansk resumé

Nyttevurderinger af transportinfrastruktur er stærkt afhængige af prognoser for både trafikal udvikling og anlægsudgifter. Disse prognoser bruges til at evaluere de forventede samfundsøkonomiske og miljømæssige konsekvenser af projekterne. Sådanne projekter er typisk kolossale investeringer, og vurderingerne spiller således en vigtig rolle i beslutninger om hvilke projekter, der er samfundsøkonomisk fornuftige at implementere, og i hvilken rækkefølge udvalgte projekter bør prioriteres. I denne kontekst er nøjagtigheden af prognoser for trafik og anlægsomkostninger vitale for validiteten og pålideligheden af den beslutningsstøtte, der forberedes til politikere. Grundet den naturlige usikkerhed involveret i at spå om fremtiden er disse vurderinger forbundet med en vis grad af unøjagtighed. Planlæggere og politikere må derfor være fuldt informeret om de usikkerheder, der er forbundet med vurderingerne de præsenteres for, for at drage fuld nytte af dem som beslutningsstøtte.

Dette studie undersøger nøjagtigheden af de prognoser, der indgår i nyttevurderinger for transportinfrastruktur, samt hvordan usikkerheden af resultaterne kommunikeres i planlægningsprocessen. Undersøgelsen leverer et vigtigt bidrag til brugen af projektevalueringsmetoder og studier i planlægningspraksis. Sammenlignet med tidligere studier af unøjagtighed i prognoser viser resultaterne af dette studie nye mønstre af unøjagtighed mellem forskellige projekttyper. Dette inkluderer vigtige skævvridninger, der tidligere har været overset, og som sætter spørgsmålstegn ved nogle af de fundamentale antagelser i vurderinger af transportprojekter. Resultaterne viser også, at usikkerheder typisk nedtones eller ignoreres i den beslutningsstøtte, der præsenteres for politikere. Denne forsømmelse får vurderingerne til at fremstå mere nøjagtige end berettiget, hvilket skaber mistro omkring resultaterne blandt beslutningstagere. Hertil kommer, at mange beslutningstagere føler, at vurderingerne kun i ringe grad afspejler vigtige politiske prioriteringer, hvilket yderligere svækker tilliden til resultaterne. Konsekvensen er, at vurderinger ofte bruges til at retfærdiggøre beslutninger taget på forhånd, snarere end at fungere som støtte til selve beslutningen.

Tre primære tilgange til at forbedre validiteten og pålideligheden af beslutningsstøtte til transportprojekter identificeres i denne afhandling. For det første skal transparensen øges i den dokumentation, der udarbejdes igennem beslutningsprocessen. På nuværende tidspunkt er det alt for besværligt at tilgå information om modelspecifikationer, nøgleantagelser og datakilder. Dette gør det svært at granske vurderingerne for eksterne aktører, hvilket skaber en black-boxing effekt, der reducerer tilliden til modelresultater. For det andet skal der oprettes et systematisk efterevalueringsprogram. Dette vil medvirke til en bedre læringsproces baseret på erfaringer fra tidligere projekter, samt levere en stærk database som både forskere og planlæggere kan drage nytte af. For det tredje skal de modeller, der anvendes i tidlige stadier af planlægningsprocessen, fokusere mere på pædagogik end præcision, da usikkerhedsniveauet er før højt til at en detaljeret analyse kan bidrage med yderligere information på dette tidspunkt. Dette vil desuden reducere black-boxing aspektet og forhåbentlig hjælpe med til at undgå nogle af de ulogiske antagelser, der eksisterer i den nuværende modelpraksis.

Abstract

Impact appraisals for transport infrastructure projects are greatly dependent on forecasts of travel demand and construction costs. These forecasts are used to evaluate the likely economic and environmental consequences of constructing new transport infrastructure. Such projects are typically colossal investments, and the appraisals thus become important both in deciding whether to implement the projects at all and in prioritising between projects selected for implementation. In this context, the accuracy of demand and cost forecasts is crucial to the validity and reliability of the decision support prepared for policy makers. Due to the inherent uncertainty involved in trying to predict the future, these impact appraisals are naturally associated with some degree of uncertainty. Planners and policy makers thus need full awareness of the uncertainties related to the appraisals they are presented with, in order to make proper use of them as decision support.

This study investigates the accuracy of the forecasts used in decision support for transport infrastructure projects and how the uncertainty of the results is communicated in the planning process. The work makes an important contribution to the fields of project evaluation methods and studies of planning practice. Compared to earlier studies of forecasting inaccuracy, the findings reveal new patterns of inaccuracy among various types of infrastructure projects. This includes important biases that have previously been overlooked, which call into question some of the fundamental assumptions used in transport project evaluation. The findings also show that uncertainties are often toned down or ignored in the decision support prepared for policy makers. This neglect makes impact appraisals appear more accurate than warranted, which causes distrust towards the results among policy makers. In addition, many policy makers consider the appraisals unable to reflect important political priorities, thereby further reducing trust in results. The consequence is that appraisals are often used to justify decisions rather than support them.

Three primary approaches to improving the validity and reliability of decision support for transport infrastructure projects are identified. First, the transparency of documentation in the decision process needs to be improved. Currently it is much too difficult to obtain information about model specifications, key assumptions, and data sources. This makes it difficult to subject appraisals to critical scrutiny, and causes a block-boxing effect that creates distrust towards results. Second, a systematic ex-post evaluation programme needs to be established. This would facilitate a better learning process from experiences in past projects, as well as provide a rich database that both researchers and planners could benefit from. Third, the models used in initial phases of the planning process should focus more on pedagogy than precision, since the level of uncertainty is too high for detailed analyses to be of much use at this point. This would further reduce the black-boxing aspect and hopefully help avoid some of the illogical assumptions that exist in current modelling practice.

1 Introduction

The greatest challenge to any thinker is stating the problem in a way that will allow a solution.

- Bertrand Russell

This chapter introduces the overall topic of uncertainties in transport project evaluation with which the present thesis is engaged. Expressed in a single sentence, the main topic relates to the validity and reliability of ex-ante appraisals for transport infrastructure projects as they are used in decision-making processes. The chapter starts off by describing the motivation for studying how, and on the basis of what knowledge, decisions on transport infrastructure investments are made, and why studies of this kind are important. It then goes on to specify the dominant forms of appraisal techniques in contemporary transport project evaluation, and the key parameters that influence their outcome. Summing up the chapter is a list of the specific units of analysis that form the core research focus of the present thesis, and which are used to identify suitable units of observation and methods of investigation in later chapters.

1.1 Transport planning and the need for appraisal

Mobility is one of the most fundamental aspects of modern society. It is probably difficult for most people to envisage how the world would look like without the ability to transport people and goods to the extent that is being done today. Food products in our local grocery stores are imported from all over the globe. People frequently travel between continents for both vacation and business. This requires that infrastructure is in place to facilitate these transport activities, whether they involve people or goods. As a result, large amounts of resources are dedicated to maintaining and improving the infrastructure that facilitates this transport, as well as developing new and innovative ways in which we can best make use of the available network capacity. For policy makers this creates an important task of selecting the most appropriate ways to deal with these issues; for planners it creates an important task of providing the most suitable analyses to guide these decisions. However, it has been observed that assumptions and forecasts in ex-ante appraisal for transport infrastructure projects are often far off target, of which ample examples shall be provided in later chapters. Time and time again we witness estimates for construction costs and traffic volumes that are out of touch with actual development. In its most condensed form, this discrepancy between expected and actual outcomes is the main focus of the present thesis, as well as how to deal with the uncertainties that causes these inaccuracies. The key aims of the analyses presented herein are to evaluate the extent of this problem in a given case area, identify conditions for when it is likely to occur, and offer guidelines to planners and policy makers on how to make better use of decision support that is plagued by inaccuracy.

Attempting to predict how future events might unfold, and what course of action would best accommodate our common needs under such expected conditions, is by no means an approach that

has been adopted only in more recent times. Written evidence teaches us that such practice has been popular throughout most of recorded history, with perhaps the most prominent example of ancient times being the oracle of Delphi. Greek and Roman leaders consulted the oracle prior to most major undertakings, allowing it to exercise great influence over political leaders in matters ranging from warfare to economics. Were the prophecies of dire nature to those seeking advice, the oracle could cause them great troubles due to the esteem in which she and her prophecies were held. When Emperor Nero consulted the oracle, he is reported to have paid great sums of money to have her prophesize as he wished (Dempsey 2003), likely to avoid any potential problems that an undesirable prophecy could cause.

Many of the early accounts of decisions based on predictions for the future are undoubtedly rooted in mythologies similar to those that surround the oracle of Delphi. That is, the existence of something or someone capable of knowing what the future brings, and thus the possibility of seeking advice in decisions that greatly depend on such knowledge. As science has gradually replaced religion as the highest level of authority in matters of knowledge, it is only natural that we have also begun looking to science rather than religion as our primary aid in decision-making. Some of the natural sciences in particular have been quite successful in establishing models that allow for fairly accurate prediction of outcomes resulting from a chain of events. No doubt this ability to provide accurate predictions is one of the main reasons why the natural sciences are today held in high esteem. Conversely, the social sciences and humanities have long been accused of a ‘physics envy’, since accurate predictions in these fields have proven to be somewhat more problematic than in the natural sciences, with physics often exemplifying the hallmark of scientific rigor¹.

Viewed in this context, transport planning is quite difficult to place in any of these traditional categories of science. This is perhaps because planning is not actually a separate discipline in itself, but rather a field of study that is approached from a multitude of different perspectives and disciplines, both in academia and elsewhere. A non-exhaustive list would at least include disciplines such as engineering, politics, philosophy, economics, geography, sociology, and psychology. In addition to this, transport planning is heavily interrelated with other broad fields, such as project management, environmental sciences, climate change, policy-making, gender studies, social justice, construction work, and land use. This is probably not surprising when we consider the great context dependency of transport as an activity, as it is exercised by many different people under a large variety of both spatial and temporal conditions. An immediate consequence of these acknowledgements is of course an object of study that is highly interdisciplinary. Inherent to this realization is also the fact that no single discipline has a monopoly on how to define the most important problems to address in transport planning.

If we look at the specific topic of ex-ante appraisal for transport infrastructure projects, the picture remains the same. As these projects require many different kinds of knowledge as well as impact many different aspects of society, this is also reflected in the requirements that policy makers set for

¹ I do not endorse such dichotomous distinctions between the natural and social sciences. Many social sciences have established models that have proven to be fairly good at explaining observed phenomena within their conditional framework, and many natural sciences still struggle in their predictive abilities (one can always bring up weather forecasts as an example). I shall discuss these matters in greater detail in chapter 2.

ex-ante appraisal. A new road scheme in an urban area might have part of its alignment located in a densely built residential area, where the potential impact on noise levels is of particular concern to the people living here. Local businesses might have a vested interest in their ability to attract customers traveling by car, and environmental NGOs might express worries about the disruption of nearby recreational areas. If part of the road is set to cross a small body of water, it might be necessary to include aesthetic considerations about bridge design, as well as environmental impacts of reduced water flow due to concrete pillars being placed along the riverbed. A privately owned ferry company might fear bankruptcy as a result of a reduced customer base, while the mayor might see a potential for attracting investors due to improved infrastructure services. The potential impacts of even smaller scale transport infrastructure projects are massive, and this should be reflected in the appraisal that is aimed at supporting decision-making. As a result, we find both elements from the natural sciences (e.g. construction design, materials science, acoustics) as well as the social sciences (e.g. budget financing, behavioural response, transport econometrics) to be integrated parts of the common appraisal techniques, in an attempt to provide decision makers with the most comprehensive overview of the likely consequences of their decisions. However, it is typically only the impacts that are easily quantifiable that are included, meaning that the appraisals are far from exhaustive in terms of impact assessment.

As can be expected, the larger the project, the more severe are the impacts. While they are rarely trivial matters, the effects of small scale projects are relatively simple to assess compared to those that affect large regions, entire countries, or stretch beyond the borders of a single nation. Some of the largest transport infrastructure projects have construction costs in the range of several hundred billions of euros, and their environmental, economic, and societal impacts can affect a wide range of stakeholders in multiple countries. For obvious reasons, decision makers should have an interest in knowing as much as possible about the extent of these impacts before initiating construction, in order to gauge the necessity and desirability of the project. Not only should the results of constructing the project be assessed, but also alternative solutions and the option of doing nothing should be taken into account in such assessments. It should be noted that alternative solutions do not have to be restricted to infrastructure construction, but could involve other initiatives depending on the framing of the problem at hand (e.g. road pricing, speed bumps, transit frequency, intelligent transport systems, urban densification). For smaller projects the costs might not count in billions, and the impacts probably do not extend much further than the local or regional scale. This does not make the assessment of them any less important though, as there are typically many more such projects than their larger counterparts, just as the economies affected are similarly scaled down in size. Systematic cost overruns and benefit shortfalls are thus just as important for small scale projects as for larger ones.

Acknowledging the complex, interdisciplinary nature of transport activities and infrastructure projects should seem uncontroversial to most readers. Indeed, most appraisal techniques employed today have also been developed with awareness of this issue, or perhaps even as a direct result of it. Nonetheless, as will hopefully become equally clear during the present thesis, this acknowledgement is sometimes disregarded or toned down in the planning process. This neglect has a profound influence on the way we conduct and use ex-ante appraisals for decision support, and consequently how we prioritize between different solutions to solve the problems we face. Regardless of scale,

transport infrastructure projects have quite significant impacts on the environments they are placed in. Consequently, the validity and reliability of ex-ante appraisals become crucial aspects in the facilitation of informed decisions.

1.2 Cost-benefit analyses lead the way

The complexities of most modern transport infrastructure projects and the systems they are placed in undoubtedly involve a great deal of uncertainty about their impacts. In response to this challenge, increasingly complex appraisal methodologies have been developed, in order to account for as many of the known impact categories as possible. Cost-benefit analysis (CBA) is probably the most popular method employed these days, and is used as ex-ante appraisal in most countries around the world (Bristow and Nellthorp 2000; Grant-Muller et al. 2001; Haezendonck 2007; Hayashi and Morisugi 2000; Mackie 2010; Odgaard, Kelly, and Laird 2005; Talvitie 2000; Thomopoulos, Grant-Muller, and Tight 2009; Vickerman 2000; Van Wee 2011). CBA methodology builds on the basic premise of comparing the total costs of a project with the total benefits derived from its implementation, which are calculated by assigning each impact category a monetary value per unit of impact. Other appraisal techniques are used to assess the expected unit volume in each impact category, which are then multiplied with the unit value and included in the CBA. An immediate consequence of this is that impacts must be quantified in monetary terms, and evaluation of transport projects thus becomes heavily influenced by economics compared to other disciplines. This is not in itself a major concern, as long as policy makers keep it in mind when interpreting the results of decision support based on CBA methodology.

The approach described above stems directly from a utilitarian approach to the philosophy of ethics, where the desirability of different choice options is ranked according to their expected utility value. In the case of policy-making we are chiefly concerned with the utility value of different policy initiatives for a bounded section of society. For transport projects the boundary is typically geographical in nature, as we expect the impacts of a given project to be reduced as a function of distance from its geographical location. Theoretically, this procedure allows decision makers to determine the feasibility of an investment by converting all impact measures to a single unit (money). It also allows for a priority assignment between project alternatives, to determine which solutions give the best value for public money. However, while the basic principles of a CBA are very easy to understand, the method is encumbered with a range of theoretical and practical problems, which are not at all trivial. On the theoretical side we mainly find disputes about assigning values to impacts and how to aggregate these into comparable units, while on the practical side we find a variety of disputes on how to arrive at a suitable measure for the impact categories that are included or how to translate the data into a prescriptive framework. Both are important in relation to the validity of CBA as a method of appraisal, and for many of these problems no obvious solutions exist. Both types of problems will be addressed in the present thesis, although the main focus will be on measuring the magnitude of impact.

In spite of widespread acknowledgement of the limitations in the use of CBA methodology (see e.g. Ackerman and Heinzerling 2004; Mackie 2010; Næss, Nicolaisen, and Strand 2012; Næss 2006; Salling and Bannister 2009; Van Wee 2011), it remains one of the most popular ex-ante appraisals for

transport planning. The reasons for this are many, but three explanations are generally accepted among both scholars and practitioners:

- **Ideology:** The utilitarian philosophy behind CBA methodology fits well with the current neoliberal political climate (Ackerman and Heinzerling 2004; Mackie 2010), which has clear links to the neoclassical paradigm that dominates mainstream economics (Clark 1998). As such, the theoretical foundation of CBA methodology square well with that of the dominant political ideologies in most of the developed world.
- **Simplicity:** The primary decision makers at in transport planning are publically elected politicians with a broad range of commitments, of which the transport sector is perhaps only one of many responsibilities. It is highly impractical (and probably even impossible) for them to go into great detail about the consequences of each project in a myriad of potential impact categories, and CBAs provide a set of simple benchmark parameters upon which projects can be evaluated.
- **Neutrality:** Since transport infrastructure projects have such far-reaching environmental, economic, and social consequences, CBAs offer a seemingly objective mode of comparison between different alternatives. This does not mean that CBA results are necessarily assumed to be decisive for policy-making, but rather that they are assumed to be a value-neutral assessment of likely consequences to support decision makers in their choices.

While there are likely many other reasons for the dominant presence of CBAs as ex-ante appraisal, these three seem to be the ones that are touted most often. Tracing the genealogy of the method brings us back to 19th century France, where the emergence of railways led to the development of evaluation techniques based on consumer surplus (Nakamura 2000; Van Wee 2011). As such, the underlying theoretical framework was arranged many years ago, but it wasn't until the 1960s that CBAs started to enjoy widespread usage in larger public works projects. More environmentally orientated methods such as the EIA has since been introduced, but the actual impact of these on decision-making has not always been very convincing (Jay et al. 2007). Even though EIAs are now mandatory in most developed countries, CBA results are usually included in these assessments and continue to play a dominant role in the decision-making process. Over the years CBAs have also started assigning unit values for an increasing amount of impact categories, resulting in an overlap between many of the traditional functions of an EIA. Greenhouse gas emissions, noise levels, and barrier effects are nowadays included in many CBAs for transport projects. CBAs thus often remain at the top of the food chain among ex-ante appraisals, which is probably why it continues to carry the most weight in decision-making.

As already mentioned, there is a wide range of impact factors associated with transport projects. A smaller set of these lend themselves readily to appraisal by CBAs, since some of them are either impractical to measure or assign a unit value. If we dig into the actual content of CBAs for transport projects, we will typically find that out of the impact factors that are included, only very few of them have any significant influence on the final results. On the cost side the construction costs are typically the dominant factor, and accounts for the vast majority of expenses. Accurate forecasts of expected

costs are thus crucial for CBA validity, just as cost overruns can be a significant burden on the economy, forcing other publically financed initiatives to be delayed or abandoned. On the benefit side the travel time savings (TTS) are equally dominant, typically making up three fourths or more of the total benefits for road projects (Banister 2008; Mackie, Jara-Diaz, and Fowkes 2001; Nicolaisen and Næss 2011; Næss, Nicolaisen, and Strand 2012). Since TTS is highly dependent on the expected traffic volumes, total benefits of transport projects are thus also highly dependent the accuracy of demand forecasts. This goes for both road and rail projects, with traffic volume measured in cars and ridership respectively². Accurate forecasts of traffic volumes are thus also crucial for CBA validity, just as transit services depend on accurate ridership forecasts for expected revenues from ticket sales.

For the reasons listed in the previous paragraph, we can use construction costs and traffic volumes as fair proxies of performance when we evaluate the ex-ante appraisal by comparison of expected vs. actual costs and benefits. In this respect, a noteworthy problem can be identified for the evaluation of benefits. Construction costs are by no means a trivial matter either, but as construction costs follow the construction of the project for obvious reasons, these can be somewhat accurately assessed once the project is completed (given that we have access to these figures). This is not the case for benefits, where the true effects on ridership and travel time savings can only be observed a while after project completion. Since CBAs account for benefits over the full lifespan of a project, and not only for benefits occurring immediately after the project is taken into use, we would have to wait 30-50 years before assessing the total benefits if we were to use a similar method as when assessing construction costs. This would require access to very detailed data over a long time span, as well as the ability to isolate effects caused by the project from other factors in this period. These are obviously unreasonable requirements if any practical post-auditing is to be undertaken, and we will thus have to make due with traffic volumes as a proxy for benefits, and usually only in the first year after opening. The implications of these inherent methodological limitations in the present study will be discussed more thoroughly in chapter 4, as well as their implications for how the results should be interpreted and used by researchers, planners, and policy makers.

1.3 Problem statement and research questions

Taking a point of departure in the planning context described in the previous two sections, we can move on to the specific problem statement that will be used to guide the work that is comprised in the remainder of the present thesis. This includes a set of research questions to be answered, as well as a demarcation from relevant areas of research that it has not been possible to cover in the present study. Some of these demarcations have already been indicated quite candidly in the previous section, and it should be clear that only with a specific subset of decision support for transport planning will be dealt with, namely impact appraisals for infrastructure projects. Within these, the scope is further delimited to a specific set of indicators, i.e. construction costs and traffic volumes. However, the primary focus will be on the traffic forecasts, and analysis of cost data is primarily used to supplement the main analysis of inaccuracies related to the estimation of travel demand. There are three main reasons for this. First, the part of the UNITE³ project in which I have been engaged is

² Car traffic also involves passengers, but this is typically done merely by scaling the benefits per car based on average occupancy.

³ UNCertainties In Transport project Evaluation (UNITE) research project.

structured solely around traffic forecasts, and there has thus not been a great deal of resources devoted to an analysis of costs. Second, cost overruns have been studied extensively in many fields, including transport infrastructure projects. Research in the accuracy of demand forecasts is sparser, and thus seem a more fruitful area to investigate. Three, I have no particular expertise in cost estimates, and it has therefore seemed more productive for me to focus on the transport related aspects of ex-ante appraisal. That being said, an analysis of cost estimates will be included as far as data availability allows, since much of the data will be made available through the sources required to do a valid assessment of demand estimates.

Further demarcations include the exclusion of non-infrastructure solutions to transport planning issues, such as road pricing, parking policies, transit subsidies, flexible work schedules, speed limits, telecommuting, ITS, travel demand management, etc. There are many reasons for this, but the primary is probably lack of resources to undertake such a comprehensive study. Other reasons include the fact that appraisal techniques for many of these solutions are not always as streamlined as is the case for transport infrastructure projects, just as some of them are relatively new and thus allow for only a limited amount of comparison between expectations and actual development. It has been deemed more resource efficient to put these related initiatives aside for now, and instead focus on a more thorough investigation of issues related specifically to appraisal of infrastructure projects.

Another important demarcation in the present study relates to geography, as the focus will be limited to projects in Denmark, Norway, Sweden and the UK. Once again this decision is mainly due to practical concerns, since much of the data collection requires personal communication and/or access to physical archives. Since the UNITE project and the majority of the partners involved are based in Scandinavia or the UK, this has been a natural geographical delimitation for the scope of the project. Previous studies have indicated problems with forecasts of both construction costs and traffic volumes in all of these countries, which have led to an increasing focus on these issues over the last decade. The political system, standard of living, and planning framework also have many similarities in these countries, which makes them well suited for a comparison across a range of projects to identify specific characteristics of projects with different levels of forecasting accuracy.

The following research questions were drafted to guide the study of the overall problem statement:

How valid and reliable are ex-ante appraisals for transport infrastructure projects as they are used in decision-making processes?

- 1. How well do forecasts presented in decision support for transport infrastructure projects match the actual development?**
 - a. What is the extent of forecasting inaccuracy in transport project evaluation?
 - b. Is it possible to identify characteristics of projects that are prone to forecasting inaccuracy?
 - c. What is the influence of forecasting inaccuracy on subsequent impact assessments?
- 2. How is uncertainty managed in the policy-making process?**
 - a. How are uncertainties communicated between stakeholders?
 - b. What are the typical explanations offered for inaccurate traffic forecasts?
 - c. Which obstacles do stakeholders identify in the use of traffic forecasts?

In this problem statement lies the implicit assumption that traffic forecasts are primarily influencing decision-making processes through their inclusion in subsequent appraisal methods, such as CBA or EIA as described earlier in the present chapter. This ought to be a relatively uncontroversial assumption, but it is hereby expressed explicitly so that readers might themselves consider the implications for the present study if this assumption does not hold. For the remainder of the present thesis the terms ‘travel demand’, ‘traffic volume’ and ‘traffic forecast’ will typically refer to estimated or expected traffic levels measured in annual average daily traffic (AADT) for road projects or annual average daily ridership (AADR) for rail projects, unless otherwise stated.

A final note to the problem formulation as stated above is the importance of the word ‘problem’. Mitroff and Silvers (2010) contrast problems with the predefined exercises often presented to students at high school level or earlier, where two of the defining characteristics of a problem are that they can be formulated in multiple ways and are inherently messy. I much agree with this perspective, but from such acknowledgement also come the realization that problems often become clear at the end of an inquiry rather than at the start. I’m also much in agreement with Ackoff’s (1999, 178–179) statement that *“a partial solution to a whole system of problems is better than whole solutions to each of its parts taken separately”*⁴. I interpret this position as necessitating a critical reflection of the initial problem formulation at the end of an inquiry whether one has been successful in the attempt to offer solutions to a given set of problems or not. Not as a relativistic blurring of the core issues through constant reformulation of the problems being addressed, but by useful application of the knowledge gained through systematic inquiry in order that the problems might be more clearly identified in the future. If no such knowledge is gained through an inquiry, the problem was likely not a problem at all to begin with, or the methods employed have been poorly selected. A brief overview of the overall structure of the thesis can be seen in Figure 1.

1.4 Summary

The purpose of this chapter has been to introduce readers to the overall context of transport planning and the role of transport project evaluation in the form of impact appraisals. Cost benefit analysis has been identified as the dominant form of appraisal methodology, and will thus be used as the point of departure for many of the arguments presented in the remaining chapters. Travel time savings and construction costs are the most important inputs for this approach, and we are thus chiefly concerned with the accuracy of forecasts for these two variables and how stakeholders deal with uncertainty in the decision support that are based upon them.

⁴ In this context I think of the societal problems we seek to address, and not the often mono-disciplinary challenges we need to overcome in specific inquiries. The best analogy I can think of is Rittel and Webber’s (1973) demarcation between ‘wicked’ and ‘tame’ problems.

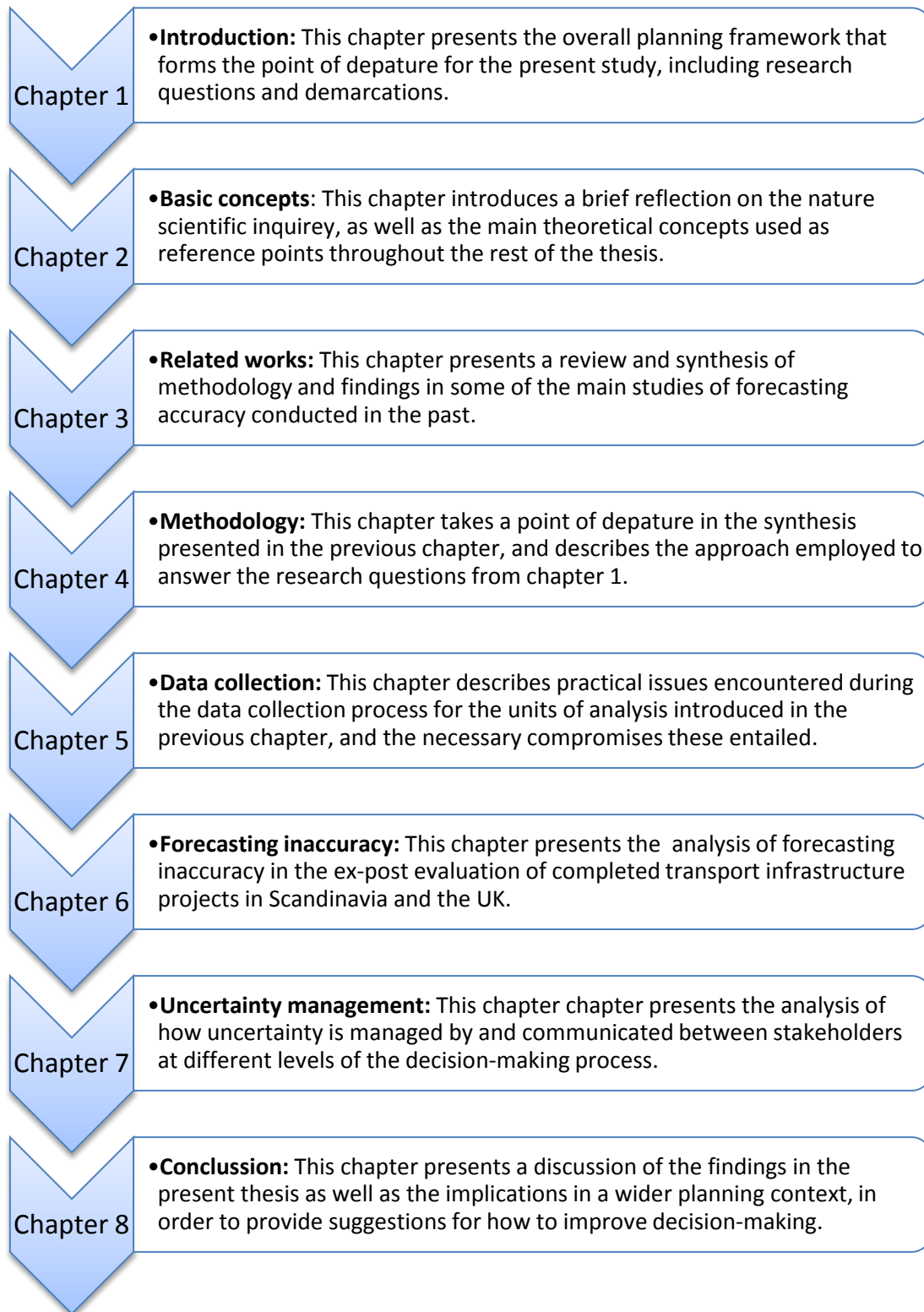


Figure 1: Overview of thesis structure and brief description of chapters.

2 Basic concepts

Science is nothing but developed perception, interpreted intent, common sense rounded out and minutely articulated.

- George Santayana

In this chapter a brief description of basic theoretical concepts related to futures studies is presented. The aim is not to provide a sophisticated display for a philosophy of science, since this is neither a primary concern of the research project nor a particular area of expertise for me. Rather, this chapter intends to discuss a set of core issues in academia, which are considered to be of great relevance to the present study, as well as research on decision-making in general. It has already been stated in chapter 1 that transport is one of the most interdisciplinary fields of research that one can engage with, and as such it is also associated with an equally broad array of methods to acquire and analyse data in the related disciplines, of which some are better suited for the specific task of appraising the value of decision support. However, in order for such methodological considerations to be of much value, they require both ontological and epistemological groundings, which is what the present chapter aims at addressing. If nothing else, the chapter will hopefully make readers able to better position themselves in relation to the arguments presented in later chapters.

2.1 Causality in a stratified reality

The physics envy mentioned in chapter 1 can be considered a response to the ability of physicist to employ their knowledge of causal effects not only to explain observed phenomena, but also to predict unobserved phenomena with great accuracy. In order to allow for such prediction it is not enough for scientific inquiry to merely describe the events that are observed. Theories must be formed about the causal relationships that explain why and when these events come about, and why in other cases they do not. When measured against their predictive capabilities, theories from physics are typically much more accurate than what we find in the social sciences. However, this is not to say that social sciences are not able to make predictive claims successfully. If that was truly the case, there would be little reason to be concerned with planning at all, since the future development of society would be completely unpredictable. In such a world it wouldn't matter much what outcome we anticipated in preparation for future conditions, since any outcome would have to be considered just as likely as any other. The entire purpose of the planning profession rests on the premise that we are able to understand social systems well enough to anticipate future development in them, at least to some minimal degree of certainty. We do so in order to identify options of intervention in cases where we are dissatisfied with the expected outcomes of current development trends. Granted, the accuracy of our predictions in the social sciences might not be comparable to those of physics, as the overall topic of the present thesis clearly indicates. There are good reasons not to expect such accuracy either, as we shall see in the following sections.

To explain why this is so, I adopt a framework from critical realism as described in Danermark et al. (2002), which is based on four primary concepts; structures, mechanisms, events, and tendencies. Structures can be defined as the pure objective reality we require for realist ontology to make sense. The structure of objects is what empowers them with the ability to exercise causal influences on other objects, hence the notion of causality. These powers are exercised through mechanisms associated with certain structures, with the mechanisms being determined by the underlying structures. The structure-mechanism relation is the nature of an object. The powers exercised by mechanisms ultimately result in events, of which some are observable and others are not (by limitation of our senses). Collier (1994) uses a wooden match as an example to describe this relationship. It is an object that has a certain structure, which determines the causal powers it is able to exercise through certain mechanisms. In this case, the typical mechanism we are interested in is the one we trigger by creating friction between the phosphorous head of the match and suitable frictional surface in order to create fire. However, we don't observe the internal structure of the match, nor we observe the mechanism that generates fire. We simply observe that a fire lights when we strike the match against a suitable surface.

This above example is intended to illustrate part of the stratification of reality that is central to this section; events are observable, but mechanisms and structures are not. Reality is stratified in a way that allows us to observe only part of it. However, events usually aren't mono causal, nor do events always trigger under seemingly similar circumstances. If we return to the match example, we can light the match in many other ways than striking it (e.g. by heating it) and we might sometimes strike it without causing a fire to light (e.g. if it is wet). In both cases the explanations are fairly straightforward, since we are now well acquainted with the mechanisms that cause matches to light. Still, this is only because rigorous testing with ignition under diverse conditions has led us to conclude that fire requires oxygen, heat, and fuel in order to be sustained. All of these three might be present without fire being caused, if other mechanisms are counteracting the ones that cause fire. We need only think again of a wet match, where we still might have oxygen, heat, and fuel present without causing a fire.

We are thus forced to infer our knowledge of the causes of ignition based on the many different observations we have of occurrences under different conditions. However, increasing our range of observed conditions is of course not only a matter of trying to ignite many different substances. Better tools have also become available to allow observations that we previously had no access to. In this case the modern theory of chemical elements, and later on the atomic theory from physics, which have led to new concepts that allowed for the modern interpretation of ignition. These theories were facilitated by advances in our ability to observe events in further detail, and not by direct observation of causal mechanisms. When Joseph Priestley isolated oxygen from mercuric oxide in the late 18th century, he did not actually observe a structure of oxygen or the mechanisms that separated it from the mercury. Rather, he observed that a new gas had been liberated and that it caused candles to burn brighter than when they were exposed to normal air. When John Dalton formulated the early versions of what became modern atomic theory in the early 19th century, he did not actually observe any atoms directly (and if he had done so he would still not be observing their mechanisms). Rather, he inferred the existence of atoms as an explanation for observable events in chemistry; in this case the formation of a new compound from two elements. Without these theories

we would be without many of the concepts we now employ to explain the cause of ignition. Earlier examples of explanations include the phlogiston theory from the 17th century, which suggested that an element called phlogiston was contained in combustible bodies, and that this gave them the ability to ignite. In fact, Priestley did not name his discovery oxygen at all, but instead called it dephlogisticated air, since phlogiston was the available concept to describe combustion at the time⁵.

Today we might think that a phlogiston theory sounds irrational, but it was perfectly acceptable given the observations at the time. Phlogiston was believed to be contained in combustible objects, and released into the air when ignited. Since an enclosed area limited the combustion of objects it was believed that a certain amount of air could only absorb a certain amount of phlogiston. This phlogisticated air could not support life, which explained observed deaths in proximity of fire but where victims had suffered no burns. The modern concept of oxygen acts completely opposite to this explanation, but it was not until people started noting that certain objects gained weight after combustion that scientists realized that the theory had major flaws. If phlogiston was indeed released during combustion, the weight of an object should decrease rather than increase. Today the theory is obsolete, but it is a fine example of how we depend on theoretical concepts to explain observable events. Even if the mechanisms they describe turn out to be wrong, the concepts might still prove useful to make sense of why certain events occur under certain conditions. We still don't know if modern theories such as that of atoms are true, but the concepts allow us to explain a wide range of observable events. We therefore trust that it is a good explanation of the underlying structure of reality, until possibly faced with new interpretations or contradicting observations that lead us to believe otherwise. The relationship between structure, mechanisms, and events are illustrated in Figure 2. Events are positioned in the concrete, observable layer in the top, and there is an external relationship with the mechanisms that govern their behaviour. Structures are positioned in the abstract, hidden layer at the bottom, and there is an internal relationship with the mechanisms they determine.

So far we have covered three of the four primary concepts I set out to discuss from Danermark et al. (2002). The fourth and last concept is that of tendencies. We have actually already been introduced to it without using the term tendencies as a reference. If we recall the match example from earlier, we observe that there is a tendency for matches to light on fire when struck, but that this is not always the case. Tendencies thus describe the events that we typically observe when trying to trigger a set of mechanisms under given circumstances. Given a strong tendency for certain events, we form a theory to explain what is causing the mechanisms to trigger these events. This was the case for the phlogiston theory, just as it was the case for its replacement. Unlike the determining internal relationship between structures and mechanisms, the external relationship between mechanisms and events is contingent, and so mechanisms do not always trigger when we attempt to make them. Another possibility is that they do in fact trigger, but are countered by other mechanisms that are

⁵ I have not included specific references for many of the claims made in this section, since they are often rather uncontroversial accounts from the history of scientific discovery. Most of the specific cases that are included here serve as illustrative examples that could be based on a myriad of other cases, and I mostly pick these at random from memory. As a result there might be some historic inaccuracies, and readers who wish to explore the development of atomic theory and the abandonment of phlogiston theory further can consult Van Melsen (1952) for specific accounts.

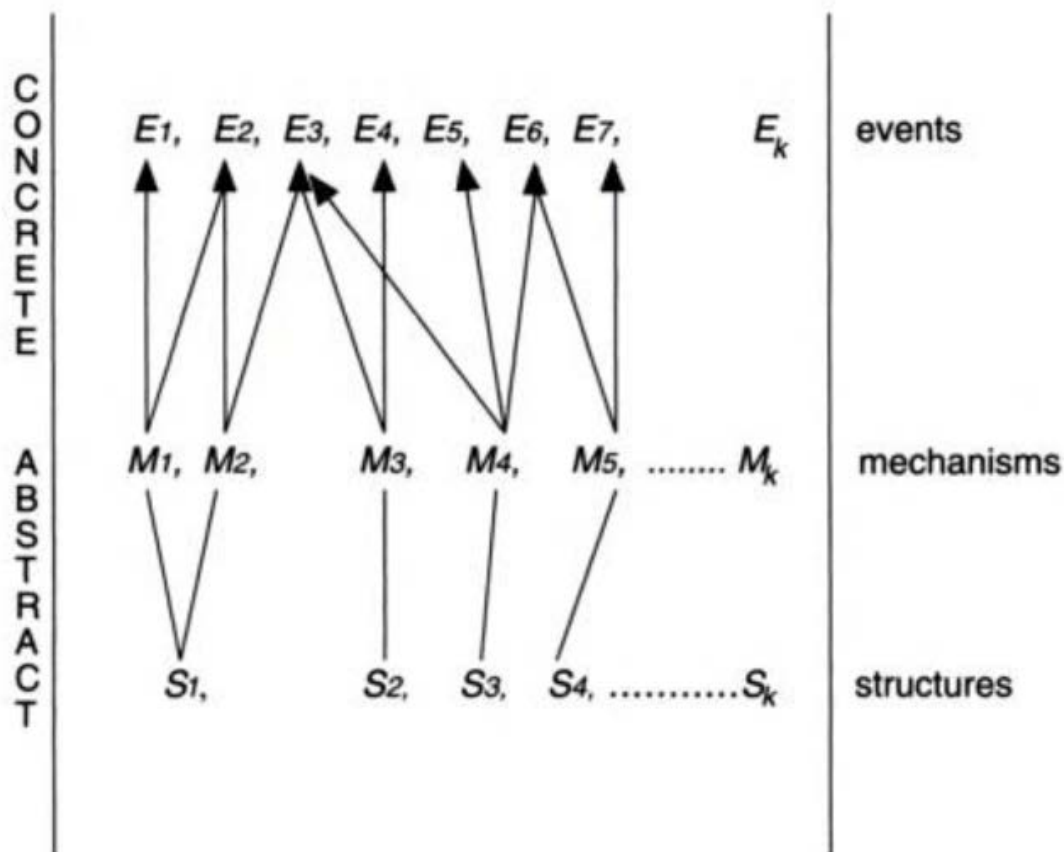


Figure 2: The stratified layer of reality split into structures, mechanisms, and events. Reprint from Sayer (1992).

also triggered, resulting in no observable events taking place. Whatever conditions are in effect when we attempt to trigger a mechanism determines both whether a mechanism will in fact trigger as well as the resulting events that we might observe. Scientific inquiry can thus not limit itself merely to establishing descriptions of observed tendencies for different events, but must constantly seek to probe the conditions that are assumed to cause the event and seek to establish a coherent theory of what mechanisms are causing it.

It is soon time to leave physics alone for a while and concentrate on the social sciences instead, but the basic premises are no different than those that have been described in the examples from physics so far. However, society is typically very difficult to confine into closed systems with well-defined conditions, since it cannot exist independently of social actors. The examples from physics thus serve as a much better introduction to the concepts in Figure 2 than those of social sciences, simply because their relationships are much simpler to grasp for most people. The main difference between the social and natural sciences comes from the stratification of reality and the emergent powers this entails. Not that stratification is unique to the social sciences, but the implications are probably more severe in this context. Danermark et al. (2002) provide a more thorough account of emergent properties and their implications for critical realist philosophy for those interested, and it shall only be covered in brief here. As an example, consider the theory of chemical elements and the atomic theory that were introduced earlier. We can reduce chemical elements to its constituents from

lower strata, e.g. in the form of atoms, but we cannot conversely reduce atoms to chemical elements. Can the objects then be reduced to the lowest known stratum and allow explanations of reality based only on inferences from observations here? Such a naturalistic claim seems quite implausible due to emergence, which is equitable with Mill's (1882, 459) concept of heteropathic effects, described as follows:

"The chemical combination of two substances produces, as is well known, a third substance with properties different from those of either of the two substances separately, or of both of them taken together. Not a trace of the properties of hydrogen or of oxygen is observable in those of their compound, water. The taste of sugar of lead is not the sum of the tastes of its component elements, acetic acid and lead or its oxide; nor is the color of blue vitriol a mixture of the colors of sulphuric acid and copper."

If we translate this into the concepts introduced by Danermark et al. (2002), Mill is basically claiming that we cannot always study the mechanisms of objects at one stratum solely by studying mechanisms of objects at a lower stratum, since new properties emerge at higher strata. Sayer (1992, 119) gives a similar account:

We would not try to explain the power of people to think by reference to the cells that constitute them, as if the cells possessed this power too. Nor would we explain the power of water to extinguish fire by deriving it from the power of its constituents, for oxygen and hydrogen are highly inflammable. In both cases objects are said to have 'emergent powers', that is, powers or liabilities which cannot be reduced to those of their constituents. This phenomenon suggests that the world is not merely differentiated but stratified; the powers of water exist at a different stratum than those of hydrogen or oxygen.

Moving into higher strata it is not always possible to trace the causal mechanisms into their constituent parts from lower strata. Since social sciences are concerned with much higher strata than natural sciences, this stratification necessarily invites for many more potential mechanisms to exercise influence over the events we observe here. Each step up the ladder adds a possibility for new powers to emerge and thus new mechanisms to account for. As such it becomes difficult, if not impossible, for researchers in the social sciences to create closed systems where they can control a desired amount of conditions. There are simply far too many potential mechanisms to control for. Researchers in the natural sciences (physics in particular) are better able to construct systems that have a desirable degree of closure, so that individual mechanisms might better be isolated. This is because they operate at a lower stratum, and thus are faced with fewer mechanisms to control for. As a result, experiments have become fundamental to advances in modern physics, while they are often highly problematic (and possibly unethical) undertakings in the social sciences. That being said, the mechanisms studied in physics are obviously not entirely reducible to closed systems, nor is it always trivial to construct appropriate closed systems under which to study them. It is merely simpler than in the social sciences, where researchers often have to rely on accounts of observable events they have little control over. This of course makes it difficult to obtain the desirable data for a given set of circumstances, since this often requires these circumstances to have been available for a number of naturally occurring events where data records are also available.

To sum up this section, I shall claim that the concept of tendencies, and its importance for our ability to form knowledge of mechanisms at higher strata, is important in relation to the degree of predictability one can expect in the social sciences, where many mechanisms are often active at any single time. Observable events are results of interactions between many different mechanisms that might negate or reinforce each other. Since we cannot observe anything but the events themselves, we are not ever able to have knowledge of the exact mechanisms that govern these events. The best alternative is then to observe a tendency for objects to behave in a certain way, and use the observations to explore the various conditions under which it does indeed behave in this manner, so that we might approximate the underlying mechanisms to form a coherent theory. There are countless conditions under which a wooden match would not light on fire for example, but typically most of them do not apply to situations where we would be in need of striking a match to light fire (and those that do are perhaps often trivial to control for). We therefore observe a strong tendency for matches to light when we do in fact strike them. The same principle applies to objects in the social sciences, but to further complicate matters we are faced with many more potential mechanisms than in the natural sciences due to stratification. The trivial conclusion here is that we should not expect predictability in the social sciences to be comparable to that of natural sciences. A more interesting conclusion is that we cannot simply infer causality by observing past events, if there is reason to suspect that fundamental tendencies might change in the future. It might even be desirable to invoke some of these changes intentionally.

2.2 Social science and models

Having covered the most basic concepts, it is now time to consider the implications for transport planning that these acknowledgements have. The first point I would like to make clear in this context is that while there are many disciplines from both natural and social sciences involved in transport planning, the specific task of forecasting traffic volumes is exclusively a matter of social science. This ought to be no controversial claim, since traffic as we understand it in relation to passenger transport is ultimately a social activity, as it is dependent on human interaction to materialize. Cars or trains do not decide to relocate at regular intervals by themselves, but are vehicles we employ to conduct various activities during the course of a day. Additionally, transport is usually considered a derived demand, since we typically do not travel for the sake of travel alone, but to reach destinations of interest so that we might undertake various activities. The task of forecasting traffic demand is thus primarily a task of understanding when and why people choose to engage in activities that require transport between different locations, as well as how they choose to transport themselves from one location to another. These mechanisms must be considered emergent properties of the social stratum, and thus we usually do not care about recent advances in DNA sequencing, Brownian motion, or quantum fields when we study travel behaviour. We are instead concerned with explaining transport related decisions via approximations to causal mechanisms, and we base these on observable events (or lack thereof) at the stratum of individuals or groups of individuals rather than the cells, elements, or atoms that constitute them.

In order to make a structured analysis of the likely responses to system changes (e.g. due to policy initiatives) we therefore construct models that attempt to explain the mechanisms we consider relevant. In this context I define a model as any simplified representation of reality that aims at providing an overview of a complex system. Figure 3 illustrates the relationship between a typical

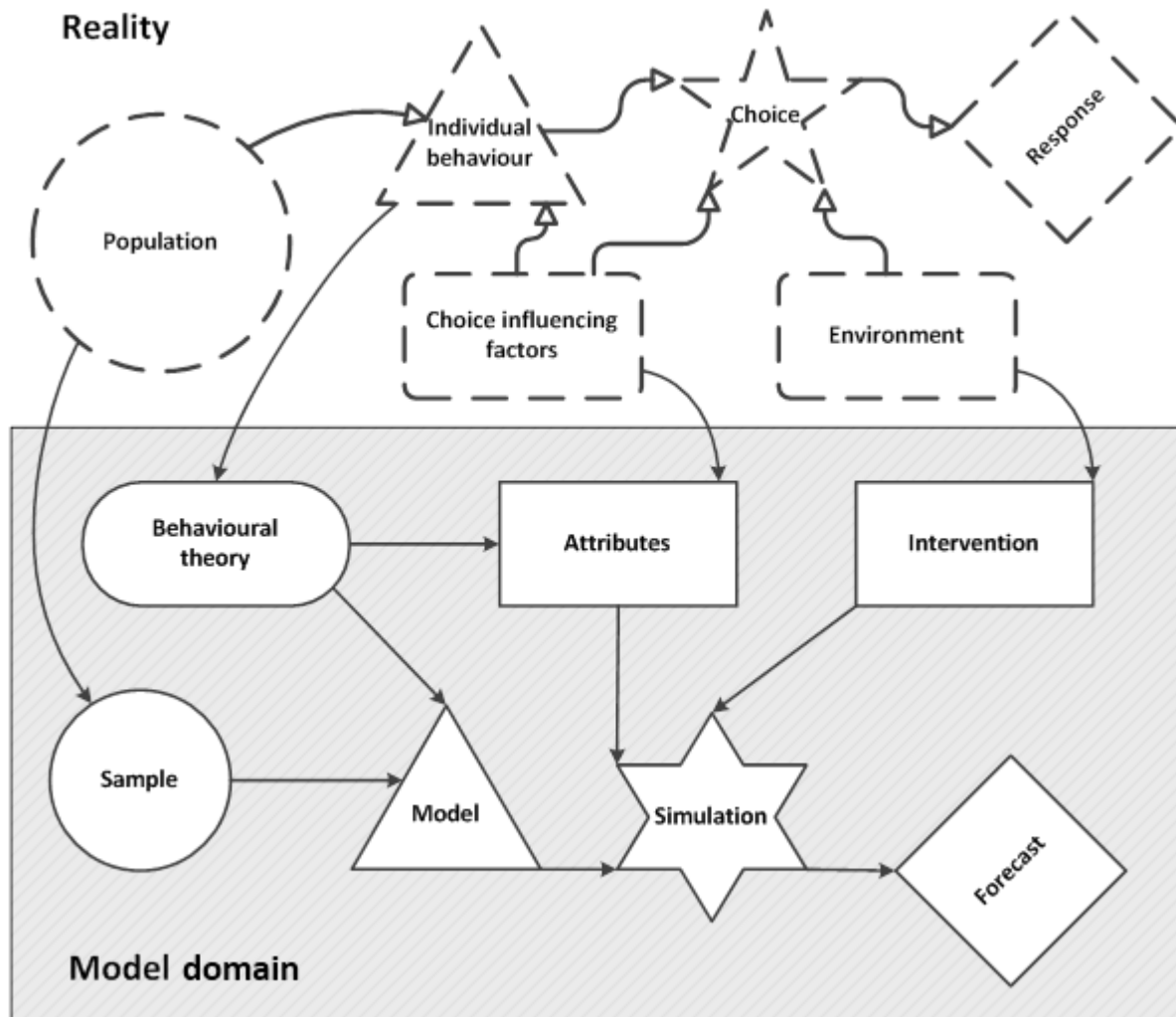


Figure 3: An illustration of a model as simplifications of reality.

model used for decision support and reality as we perceive it, and the illustration is thus in itself a conceptual model. This definition includes both conceptual and mathematical models, where the former focuses mainly on pedagogy while the latter focuses mainly on precision. There are many ways to go about modelling system responses to interventions, but in relation to transport models the traditional four stage approach has been quite dominant. I shall elaborate on the implications of this in section 2.5.

Let us look a bit more in detail at the processes in which models usually play a prominent role. Walker et al. (2003) present policy-making as a multistage iterative process that serves as a good starting point for discussing this process. Figure 4 displays a slightly modified version of the process described by Walker et al. that I feel highlights some of the issues relevant to the present study better than the original version. It should be noted that decision-making in practise is sometimes quite different than the harmonious process being described here, and the structure of the process in Figure 4 is probably more normative than descriptive. The general decision-making process starts at the top of the loop (1) with the identification and framing of a problem as a result of dialogue among policymakers, stakeholders, and analysis. The outcomes of interest from this stage are then

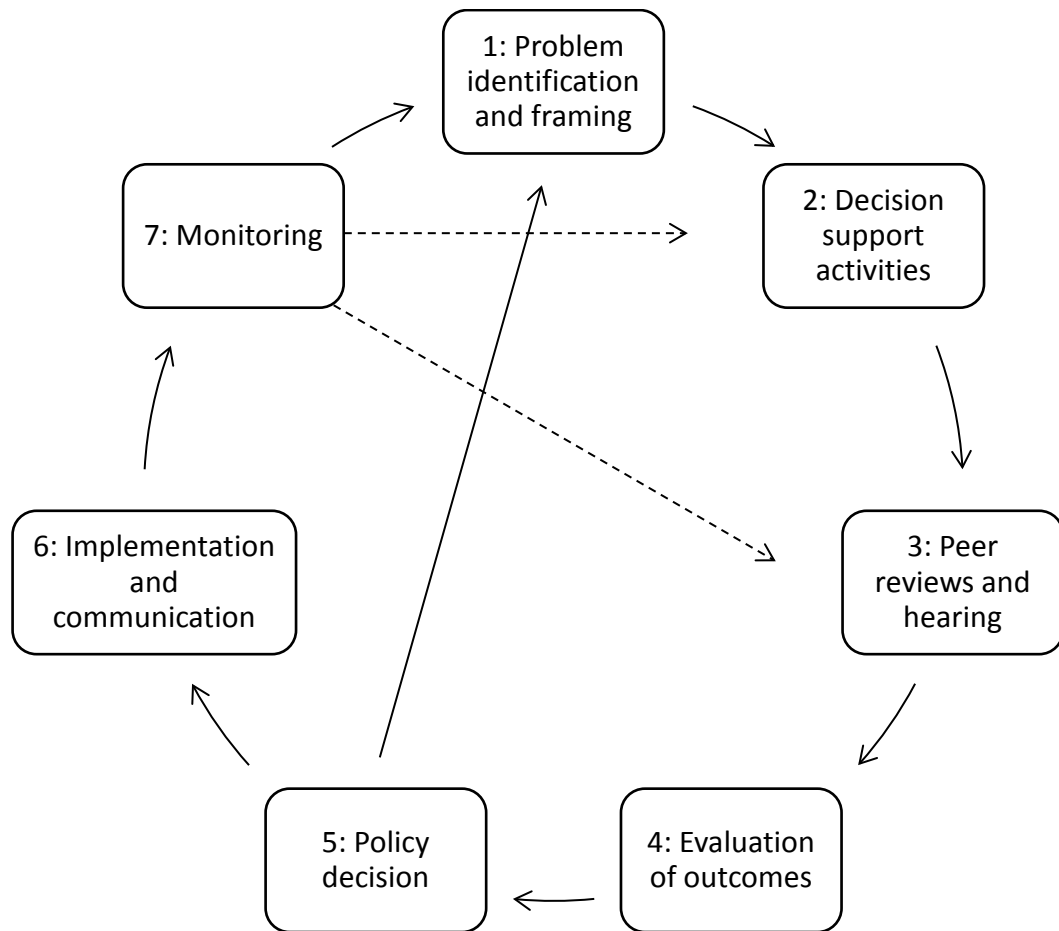


Figure 4: The policy-making process as a multi-stage iterative process. Inspired by Walker et al. (2003).

assessed by policy analysts (2) that prepare the formal decision support, and this is typically the phase of the model that relies on mathematical modelling the most. In the peer review stage (3) the results of the decision support is scrutinized by experts and subject to public hearing (this is the case for most public works projects at least). The decision support is then evaluated by policy makers and stakeholders (4) based both on the results from phase 2 and the critique from the peer review process and public hearing sessions. A decision can then be formed on how to proceed (5), and if it is found that the prepared decision support is inadequately scoped, highly uncertain, based on partisan bias, or otherwise unsuited for decision support, the process is returned to stage 1. Otherwise the chosen alternative is implemented (6). The final stage is the monitoring of impacts to evaluate whether the objectives from stage 1 are being met, and whether uncertainties from stage 2 have been resolved. I have added a dashed line from stage 7 to stages 2 and 3 to indicate that this monitoring provides crucial knowledge for the preparation of future decision support as well as the peer review by experts.

The process presented in Figure 4 is intended to serve as a generic conceptual model for policy-making, if we define a policy as an initiative by administrators of a system in order to change the way it operates in order to obtain a specific outcome. The motivation for such initiatives can of course be quite diverse, but they are usually focused on solving conflicts or obtaining benefits. If we frame transport planning in this conceptual model, the main objectives typically include faster travel time,

congestion relief, accident reduction, noise reduction, improved air quality, or increased mode share for public and non-motorized transport. The model-based decision support in stage 2 is where transport models are most relevant, and it is also forecasts produced at this stage we are interested in when comparing expected and actual traffic volumes. This comparison can be considered a type of monitoring at stage 7 for a small subset of the parameters in the analysis. Due to the complexity of the system we wish to study, the models are often considered an indispensable tool when preparing the decision support. Conceptual models can aid us in this, and we use them to describe the layout of the system as we envision it. Conceptual models are especially useful to communicate the structure of a system in a pedagogical manner, and are good at facilitating a reference point for discussing the mechanisms in a system. For example, the process in Figure 4 is presented via a conceptual model, that (hopefully) makes it easy for readers to understand how I envisage policy-making, and enable them to evaluate whether I have left out important mechanisms (which is undoubtedly the case).

However, conceptual models are rarely enough to satisfy the needs of decision makers, who often require inputs that can transfer into subsequent impact assessments. For this we need mathematical models to produce numerical outputs. This is especially true in transport planning, since many of the subsequent impact assessments are highly dependent on numerical inputs (e.g. EIAs or CBAs). Whereas a conceptual model can be as simple as a few descriptive boxes and arrows to illustrate the interrelations between different objects, a mathematical model expresses these boxes in numerical values and the arrows as algorithmic functions. The mathematical model can thus be considered an advanced conceptual model that expresses the relationships in a way that allows us to calculate outcomes as quantities. The quantification of inputs and relationships is done by observing how people decide to travel in different circumstances, and the model is imported in a computer program to facilitate the necessary processing power. Due to the large systems being modelled it can be a time consuming process to run these computerized models, even with continued advances in processing power. It is these models that form the backbone of the decision support activities at stage 2 in Figure 4, where they are used to explore the responses of complex systems to a range of different policy initiatives. Such a role of models is illustrated in Figure 5, where we see the model envisioned as a mediator of knowledge for policymakers to help them cope with the complexity of the system context in order to set the goals, objectives, and preferences for a policy initiative in a way that results in the most desirable outcome.

It should be obvious that this sophistication from conceptual model to mathematical model involves many assumptions to be made along the way, of which only a limited amount can be based on substantial observations of events in the individual case. When a model framework is institutionalized, as has long been the case in transport planning, the models at individual steps often become very standardized. Each model sits in a complex relationship with other models, upon which it depends to quantify all of the necessary inputs and parameter estimations. A transport model requires assumptions to be made about what the relevant influential parameters are, which of them can be reasonably estimated, how sensitive the system is to changes in them, etc. It would be a daunting task to go through this laborious process every time a model was to be constructed for a specific case, and so many of these assumptions are settled by the use of generalized quantifications

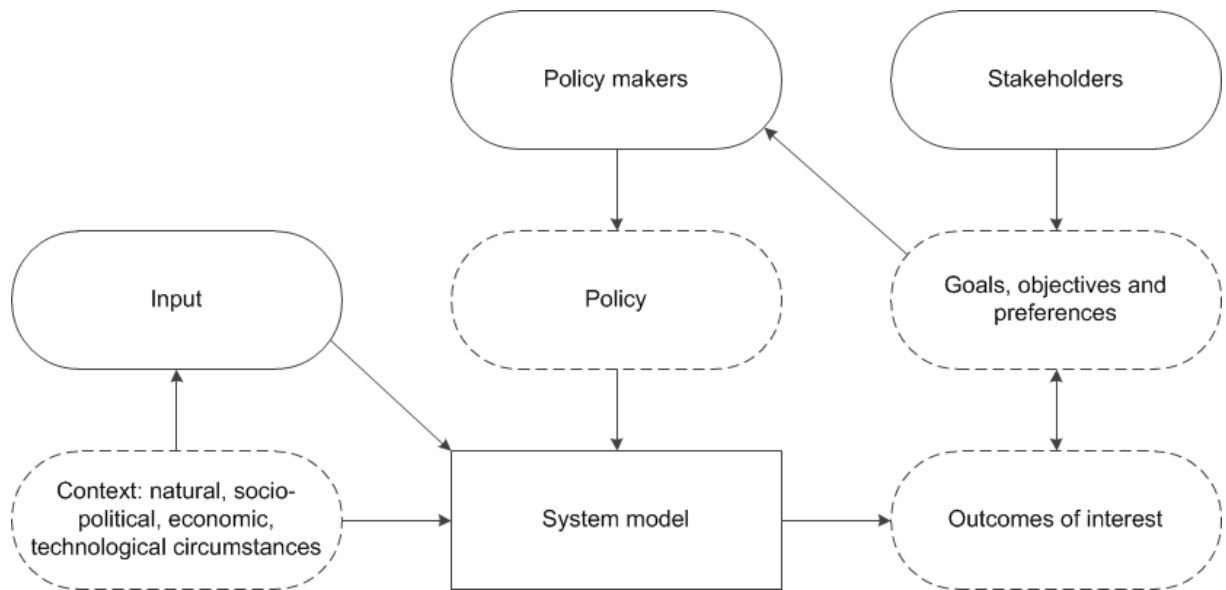


Figure 5: The role of the system model within the policymaking process. Based on Walker et al. (2003).

of these relationships. The same arguments can be put forward for the subsequent impact assessments, just as many of the input variables that are fed to the transport model are estimated by standardized assumptions from other disciplines (e.g. population, income, car ownership, etc.). The result is a generalized appraisal tool that might misrepresent the specific system of an individual case to a certain extent, but which is considered a reasonable representation of how such systems behave in general.

Modellers are obviously aware of many the above generalizations, and no text-book on transport models claim these tools to be crystal balls able to predict the future. Forecasts are often described as qualified guesses bound up on a systematic analysis of relationships, where external variables are typically evaluated based on scenarios. A scenario in this regard is understood as a consistent set of assumptions, and the purpose of scenarios is thus to check how the system performs under a variety of such assumptions since we don't know which of them will hold true in the future (this is true for both input variables and model specification). The construction of scenarios is often more of an art than a science, since observations of short term trends could easily lead to erroneous conclusions about the future if they are only a result of temporary disruptions in the system, while failure to account for them could be crucial if they are the indication of a major societal change. In the former case modellers risk overemphasizing the influence of transitory fluctuations, where an example could be extrapolation of oil prices during the crisis in the 1970s. In the latter case they risk failure in providing decision support on important system changes, where an example could be the rapid introduction of women in the workforce after World War II. The temporal uncertainty is especially problematic since it might cause dramatic changes in the external relationships between mechanisms and events (cf. Figure 2). The point of these reflections is to highlight the importance of documenting the forecast related uncertainties that are not internal to the model or mere sensitivity analysis of the inputs used. These may be more important for the accuracy of the forecast than the model specification.

2.3 Classifying uncertainty in model-based decision support

In this section I will expand upon the general framework of uncertainty that started with the epistemological and temporal categories from the last chapter. Walker et al. (2003) suggested a framework based on three dimensions of uncertainty in model-based decision support: nature, location, and level. While the classification is more detailed than I consider necessary for the present thesis, it does provide a good point of departure in discussing uncertainties in relation to the use of transport models in transport project evaluation. In this section I will briefly describe the dimensions included in this framework, and how they might be of relevance to the analyses in later chapters.

2.3.1 Nature

Walker et al. (2003) classify the first dimension in two different categories; epistemic and variability uncertainty (or: epistemological and ontological uncertainty). They define epistemic uncertainty as *“the uncertainty due to the imperfection of our knowledge, which may be reduced by more research and empirical efforts”* and variability uncertainty as *“the uncertainty due to inherent variability, which is especially applicable in human and natural systems and concerning social, economic, and technological developments”* (Walker et al. 2003, 13). The variability category is split into subcategories of natural randomness, societal variability, and human variability. Natural randomness can be described as the chaotic nature of natural processes that we are unable to predict, with the weather being a good example. Societal variability refers to social, economic, or cultural dynamics at the macro-level, where one example could be current rise of digital communication options that seemingly reduces the car dependency of young people (Koslowski 2011; Sivak and Schoettle 2012), which could dramatically alter future requirements for transport systems. Human variability refers to the non-rational behaviour of individuals at the micro-level, where potential cognitive dissonance might be important in relation to stated and revealed preferences. What Walker et al. (2003) are really describing thus seem to be spatial and temporal uncertainties at different strata of reality, with the societal variability being an expression of the highest strata among these⁶. It is also the one that is typically the most interesting in relation to transport planning for larger transport infrastructure, although micro-level behavioural changes or natural randomness are by no means irrelevant. Consider for example the impact on air traffic by the 2010 eruption of the volcano Eyjafjallajökull, which caused severe disruptions and resulted in many requests to implement back-up systems in case of future such incidents. Issues related to climate change adaptation might also be an area that is considerably vulnerable to natural variation, even if there are severe anthropogenic causes behind many of the climate related problems we currently observe observing.

2.3.2 Location

The generic location categories suggested by Walker et al. (2003) are context, model, input, parameter, and outcome, with the last of these being the accumulated uncertainty in the outcome of the model. Context refers to the framing of the system boundaries in terms of what to model, since

⁶ I realize that some might consider human behaviour to be at higher strata, but I find this a misunderstanding of the concept as they apply to this context. Societies are made up of individuals, and as such exist at a higher stratum, just as when we distinguished between chemical elements and their constituents (atoms). Some (not all) of the mechanisms in societies can be explained by tracing them to those of individuals, but those of individuals can never be explained through reduction to those of societies. However, that does not mean that events at the stratum of societies cannot trigger the mechanisms of individuals.

stakeholders do not always agree on what the target goals for an initiative should be. Since a model is always a simplification of reality it becomes crucial to the validity of a model to clearly define the most important issues that we wish a model to address. Global warming is an example of an issue that most people consider a problem, but where there has traditionally been a high amount of disagreement on the framing of the problem. Some consider it to be mainly a technological problem of production efficiency while others call for economic regulations such as emission quotas to incorporate environmental externalities and affect rates of consumption. Others still might view it as a cultural problem rooted in our levels of affluence. All of these framings might be relevant, but since we cannot model everything, we only include those that are deemed the most important by stakeholders and policy makers (at least to the extent that modellers are able to implement them). To take an example from transport planning, we might consider lower transit fares as a measure to combat urban congestion, since we expect car users to gain an incentive to switch their preferred mode of transport for commute. However, such a measure is difficult to capture in a model that does not incorporate transit fare elasticity in its framework. While it might be possible for modellers to adjust the model to reflect this effect, such adjustments will obviously be a sub-optimal solution compared to a model that was built with the specific purpose of evaluating transit fare changes. Stakeholders need to be explicit from the start about the context in which the model is to be used as decision support.

The second location category is the uncertainty stemming from the internal structure of the model once the context has been framed. Walker et al. (2003) distinguish between structural and technical model uncertainty in this category. The structural uncertainty is described as the lack of understanding that we have of the system, since we might be unsure about how the variables we include in the model are related to each other. This is a problem mainly associated with the conceptual model, but it obviously translates into the mathematical version of it as well as the technical implementation in a computer. The technical uncertainty is described as the software or hardware errors that might give cause mistakes in the processing of data, and is therefore specifically related to the computerized version of mathematical models.

The third location category is uncertainty of the input data that is used to describe the system. Like the previous category this is split into two subcategories by the authors; external driving forces and system data. The first relates to the actual effect of the input variables that drives system changes, while the latter relates to the measures of relevant features for a specific system. For example, oil prices might change drastically, which is uncertainty related to an external driving force. Conversely, data on land use might be outdated, which is an uncertainty related to the system data. Both types of uncertainty have a considerable influence on travel demand.

The fourth location category is the parameter uncertainty. This relates to the constants that are used in the model, which are split into four subcategories; exact, fixed, a priori, and calibrated. Exact parameters are universal constants such as π . Fixed parameters are well determined constants such as gravitational acceleration. A priori parameters are less well determined but a fixed value is still used (e.g. elasticity values). Calibrated parameters are essentially unknown and need to be adjusted to the specific geographical context of the model (e.g. friction factors). It should be obvious that there is more uncertainty in the two latter subcategories than in the two former, which are of

relatively minor importance to the propagated uncertainty of travel demand models. Since transport planning is dominated by social rather than natural sciences, this should be unsurprising, as there are typically very few universally fixed parameters in social sciences. The last location category is the accumulated model outcome uncertainty that is basically the total propagated uncertainty resulting from the four previously described categories.

2.3.3 Level

Walker et al. (2003) define three levels of uncertainty between the two polar extremes of determinism and indeterminacy; statistical uncertainty, scenario uncertainty, and recognized ignorance. Statistical uncertainty refers to the inaccurate measures of true values, which can have a number of different causes. The sample population might not have the same characteristics as the population we wish to apply our system model to, it might be too small to arrive at a good estimate, or the spread might be so large that a large deviation from the mean is common. These subcategories are fairly standard in statistics, and what the authors describe are perhaps more commonly referred to as sampling error, standard error, and standard deviation respectively. The second level of uncertainty is the scenario uncertainty, which is described as the inherent uncertainty that stems from making assumptions about the future condition of the system we wish to model. Unlike statistical uncertainty it is difficult to quantify scenario uncertainty by associated probabilities, since scenarios are necessarily discrete rather than continuous. As mentioned earlier, scenarios are sets of coherent assumptions about the future and likely more of an art than a science. The price of oil is a good example of a variable that is highly influential on the cost of travel, but where the development is fairly unpredictable and subject to large fluctuations. In such a case we are forced to make do with scenarios of the likely development of oil prices, but this cannot be expressed meaningfully in statistical terms⁷. The last level of uncertainty is recognized ignorance, which is described as a situation in which we have little to no knowledge of the functional relationships and thus have no basis for scenario construction, but are at least aware that we have no knowledge of these matters. A fourth pseudo-category is one that extends into total ignorance, since we can never have a full overview of the things we are unaware that we do not know⁸.

2.3.4 Using the classification in practice

The attempt at constructing an overall framework for defining uncertainty by Walker et al. (2003) is quite thorough, and gives a good overview of the many different types of uncertainty that modellers are faced with when producing forecasts. It goes without saying that policy makers that form their decisions partly based on model-based decision support are faced with the same uncertainties as those of the models, but they are typically less aware of these uncertainties than the modellers are

⁷ Unless we assumed it to grow at a fixed background rate while allowing for some temporary fluctuations, in which case we can express the uncertainty numerically. This is obviously an unrealistic model, and only masks scenario uncertainty as statistical uncertainty by making a new set of uncertain assumptions. Sadly, it is not uncommon to see scenario uncertainties being handled this way in decision support under the guise of sensitivity analyses.

⁸ This is now also known as 'unknown unknowns', a term popularized by Donald Rumsfeld (2002) when referring to the absences of evidence for weapons of mass destruction in Iraq. However, I doubt that Walker et al. (2003) would agree with Rumsfeld that mere concern about unknown unknowns is a justifiable reason for going to war, and as it later turned out it was perhaps more a case of 'unknown knowns': a refusal to admit that intelligence had confirmed that no such weapon were likely to exist.

(Duncan 2008). However, not all of the categories identified by Walker et al. (2003) are equally relevant to the current study, and I find some of them rather difficult to operationalize for practical use in assessment of model-based decision support. This is not to say that they are not relevant categories of uncertainty for an overall framework, but the main purpose of including different dimensions of uncertainty in the present study is to have a general taxonomy of uncertainty types to use as a reference for the topic at hand, which is transport project evaluation. For example, the technical model uncertainties stemming from potential flaws in hardware and software tools must be considered a very specialized type of uncertainty, which I have little chance of investigating.

In addition, I also consider the nature dimension as an overall classification that can be applied once the location has been identified. The level of uncertainty is considered dependent on the location and nature dimensions. The overall idea is fairly similar to that of Walker et al. (2003), but I consider the categories to be guidelines for uncertainty evaluation rather than an exhaustive array of mutually exclusive types. The location dimension seems the most suitable as an initial identifier in this regard. To give an example of a classification we can consider a distribution sub-model in a traditional four stage model, which we shall return to in section 2.5. The uncertainty related to this is mainly located in the model structure as a specific representation of mechanisms, since a method of distributing trips must be specified. The nature of the uncertainty is mainly epistemic, since we know that this is an approximation of the mechanisms that govern travel behaviour no matter which model we choose. The level of uncertainty is considered a statistical risk, since it is possible to validate the results of the distribution model with observed values in the network. However, there is also a temporal dimension of uncertainty in the calibration of the distribution model, since it is validated based on past conditions. There is no guarantee that the calibrated parameters from the reference situation will reflect future conditions well. Preferences might have changed, which is perhaps best classified as scenario uncertainty or recognized ignorance. There are of course also non-statistical uncertainties such as doubt about model specification validity, but these are typically disregarded when reporting confidence intervals of results since they are not quantifiable uncertainties (if even recognized at all).

The framework from Walker et al. (2003) will be used as a general reference for uncertainties throughout the remainder of the thesis, although often with focus on only the location or level dimensions described here. The framework is mainly intended as a platform for discussing the influence of specific types of uncertainties on the model outcome, the extent to which it is possible to reduce uncertainty of these types, and the communication of them to decision makers.

2.4 Accuracy and the role of planning

Having covered the more philosophical underpinnings of the concepts employed in this thesis we shall now turn to the concepts that will form the core of the analytic approach. The concepts of accuracy, precision, and bias undoubtedly fit into this category, as it is observed experiences of these that have been the prime motivator for the present study. Accuracy can generally be considered an aggregate of precision and bias, as illustrated by the analogy of hitting targets at a shooting range (see Figure 6). Precision is considered as the observations' degree of deviation around some central node while bias is a marked tendency for observations to be located in a specific direction from this node. When faced with imprecision, the approach to improving accuracy is to limit the spread of

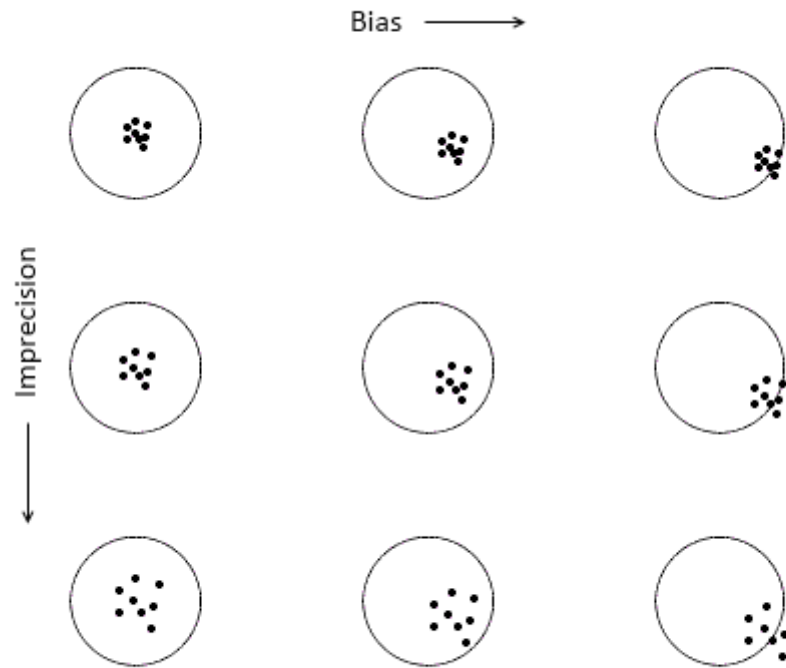


Figure 6: Lack of accuracy can be a result of imprecision and/or bias. Inspired by De Jongh (1998).

observations. When faced with bias, the approach to improving accuracy is to bring the centre of observations closer to the node.

In transport planning it is clear that a certain degree of accuracy must be obtained in forecasts if they are to be of much use. We are unlikely to achieve perfect accuracy in our predictions, especially considered the broad scope of potential uncertainties that have been described in the previous sections. However, a high degree of imprecision will reduce the ability of forecasts to serve as decision support. The ultimate purpose of travel demand forecasts is to serve as input to subsequent impact assessments such as CBAs and EIAs, but an imprecise measure makes it difficult to use these evaluations in a ranking exercise, whether it is between different proposals or in comparison with a zero-alternative. After all, if the result is very imprecise we will be likely find ourselves in situations where the ranges of possible values overlap for each project, in which case they should all be treated alike. Another option is that this imprecision is ignored, in which case we will achieve a false ranking since it hides the uncertainty in the estimates. For example, if two projects display a benefit cost ratio of 2.2 and 2.9 respectively, we would consider the latter of them the better alternative, *ceteris paribus*. However, if the imprecision of the measure means that we ought to consider a range of ± 0.4 from these nodes we find that the resulting ranges will overlap, and it thus becomes impossible to distinguish meaningfully between them. A biased measure poses a different but equally important concern, since biased result will skew the ranking of certain projects in one direction, making them appear more favourable than they ought to be. Imagine that the environmental disruption caused by construction is not included in the EIAs for new transport initiatives (this is typically the case). Since we only compare the expected impacts of the use of these facilities we will thus tend to favour

projects that include a lot of construction work, *ceteris paribus*. Both precision and bias are problematic, but eliminating either is unlikely to happen. Policy makers must therefore carefully consider how much accuracy they demand from the decision support they are presented with. Conversely, planners and modellers need to consider how the uncertainties at various point in the preparation of decision support contributes either imprecision or bias to their results, and make sure to communicate the limitations in their results.

When we speak of accuracy we are thus chiefly concerned with these two indicators. If we turn our attention to transport planning, there are of course a great deal of variables to be considered, of which each are associated with various degrees of precision and bias (even if we cannot quantify these in a meaningful way). Many of them are probably also estimated by other variables that have their own limitations on measurement accuracy, and thus uncertainty propagates through the chain of modelling. What we are chiefly concerned with in the present study is traffic flows however, and we shall therefore concentrate on accuracy measures of the amount of observed traffic compared to that of our expected values⁹. The traffic forecasts thus become our nodes and the traffic counts our observations. For a measure of inaccuracy that is fairly easy to interpret I have chosen to following definition¹⁰:

$$I_P = \frac{O - P}{P} \times 100$$

In the remainder of the present thesis inaccuracy will thus refer to values in the form of I_P , if nothing else is specifically stated. To the extent that good count data and access to decision support documents are available, we now have a reasonable measure for evaluating the accuracy of decision support. For a sufficiently large sample it should thus be possible to assess both precision and accuracy of traffic forecasts via inaccuracy measures in the form of I_P . Unfortunately, the assumptions regarding data availability are quite optimistic, as will become evident in chapter 5.

In order to assess proper degrees of accuracy for traffic forecasts we need to reflect on what the purpose of them really is and how they fit into the role of planning. If we recall Figure 4 and Figure 5 we see that transport models are primarily used as descriptions of a complex system in order to provide an overview of the expected impacts from various initiatives. Hall (2002) describes the process in the following way:

Having taken the basic decision to adopt planning and to set up a particular system, planners then formulate broad goals and identify more detailed objectives which logically follow from these goals. They then try to follow the consequences of possible courses of action which they might take, with the aid of models which simplify the operation of the system. Then they evaluate the alternatives in relation to their objectives and the resources available. Finally, they take action [...] to implement the preferred alternative. After an interval they review the state of the system [and] begin to go through the process again.

⁹ As well as costs to some degree, but the approach outlined here applies equally well for that.

¹⁰ P: predicted value, O: observed value

In transport planning we are typically concerned with consequences like safety, pollution, accessibility, noise, barrier effects, social exclusion, etc. We can group these into three main categories; economic, environmental, and social. The two first groups are largely determined by the amount of traffic that we expect, while the latter is mainly dependant on how we chose to design the infrastructure and associated regulations. The social consequences of traffic forecasts are thus less dependent on traffic volumes than the environmental and economic consequences if we consider the direct impacts on assessments. However, as prioritization between different projects is based on subsequent assessments there can of course be indirect consequences of inaccurate traffic forecasts that spill into the social sphere.

The economic and environmental consequences are more direct. A rise in road traffic means a rise in emissions, accidents, noise, and congestion. Congestion reduces travel times and vehicle efficiency, leading to further emissions and a less of efficiency. A drop in road traffic is associated with reverse effects. For rail traffic there is typically little change in environmental impacts as a result of passenger volumes, since empty trains still require fuel and produce noise. However, if passenger loads become larger than the system is able to handle there is of course a potential for severe crowding issues at platforms. Loss of efficiency is then also a problem. Conversely, empty trains result in no revenue, so the economic consequences of failed traffic forecasts should be obvious here. The role of transport planning from a forecasting perspective is then primarily to produce accurate figures of traffic, which for roads is the amount of vehicles and for rail the amount of passengers. As a further concern we are primarily interested with rush hour traffic, since there are typically no capacity problems outside this of the peak periods. This is especially important for road traffic, since congestion increases travel time, which is the primary input for cost benefit analyses. It also lowers fuel efficiency drastically, which is important in evaluating CO₂ emissions for environmental impact assessments. However, average annual daily traffic (AADT) is typically a reasonable indicator of traffic levels, as peak hour traffic is usually proportional to this measure. I shall therefore only consider AADT values when comparing forecasts with count estimates and assume that any inaccuracy in this measure will also be present for peak traffic¹¹.

2.5 Four stage models

Travel demand modelling is a complex discipline that I shall not at all attempt to cover more than very briefly in this section. The purpose is not to familiarize readers with the technical aspects of modelling, but to give an impression of the kind of limitations that arises from various approaches to the simplification of reality that is a necessary consequence of employing transport models as decision support. Transport modellers are, ideally at least, highly specialised experts that have combined the techniques of transport planning with traffic engineering. Typically they have a master's degree in engineering, as transport modelling originally developed as a tool to assist highway engineers, and has since been taught within engineering faculties at universities. There are good reasons for this, since the nature of contemporary modelling efforts is highly technical, and transport modellers might spend up to 90% of their time with programming efforts and data

¹¹ Depending on the type of project, this assumption might have rather important effects on the interpretation of results. Data for a more sophisticated analysis have not been available, but this limitation should certainly be kept in mind.

management (Rich 2010). There are of course a wide variety of models, of which some are better suited than others depending on the purpose of modelling. For example, detailed simulation models might be well suited to describe the behaviour of individual drivers in a closed network. This can be a very valuable tool for evaluating initiatives at the micro level, such as specific sections of a street map or the coordination of traffic lights. These are often highly dynamic, since feedback is an important part of modelling the response of individual drivers to changes in the system (not least the behaviour of other drivers).

However, such models are not very useful when we wish to assess the traffic volumes on large sections of a highway network 30 years ahead, or the expected impact of a new light rail system in an urban area that previously had poor transit facilities. For these purposes the typical approach has so far been to use four stage models (FSM). These provide a mechanism to determine equilibrium flows in a network, and originated in the US during the 1950s. They were imported to Europe during the 1960s (Bates 2000), and while a multitude of new approaches to modelling has since appeared, the FSM approach is still by far the dominant in practice (Bates 2000; McNally 2007; MOTOS 2007; Murga 2002). The purpose of these models was originally to aid in the provision of highway capacity, since transport planning was predominantly occupied with the challenges imposed by the rapid increase in car ownership after the Second World War. Today, the challenges that face transport planning are quite different, with congestion, CO₂-emissions, and social exclusion being important adverse effects of rising car dependency that needs to be addressed. However, there has been no radical break with the traditional approach to transport modelling (Ortúzar and Willumsen 2011), and while contemporary FSMs are more sophisticated than earlier versions, this advance has taken place in an evolutionary rather than revolutionary setting. This means there is a strong path dependency present in the evaluation of transport projects due to design of tools such as transport models. Failure to recognize this could lead to some unfortunate misunderstandings about the likely impacts of transport infrastructure projects, if these models are used to assess policy initiatives that trigger mechanisms that the models are poorly suited at describing.

A basic outline of the underlying rationality used in a four step model can be seen in Figure 7, with the main input categories on the left and the four model steps in sequential order on the right. On top of this is of course the data necessary to define travel behaviour, which is used to calibrate the parameters of the model. Typically such data is acquired through a variety of questionnaire and interview efforts as well as observed traffic flows on existing networks. It should be obvious that the data requirements for this type of modelling are quite substantial if we wish the model to be a fairly accurate reflection of reality. It is both costly and time consuming to gather the necessary data and models are therefore updated only when it is deemed cost-effective. It is not unusual for traffic forecasts to be based on models that are calibrated on the basis of data that was collected a decade or more ago, which weakens the link between model and reality since travel patterns might have changed in this period. Furthermore, forecasts are often made a decade or more into the future, which adds further temporal uncertainty to the forecasts. Travel behaviour in the model is assumed to be constant over large spans of time, while travel behaviour in reality changes all the time.

The first of the four steps is the trip generation, which aims at defining the overall volume of travel in the system. Travel behaviour is typically considered a derived demand from the activities we seek to

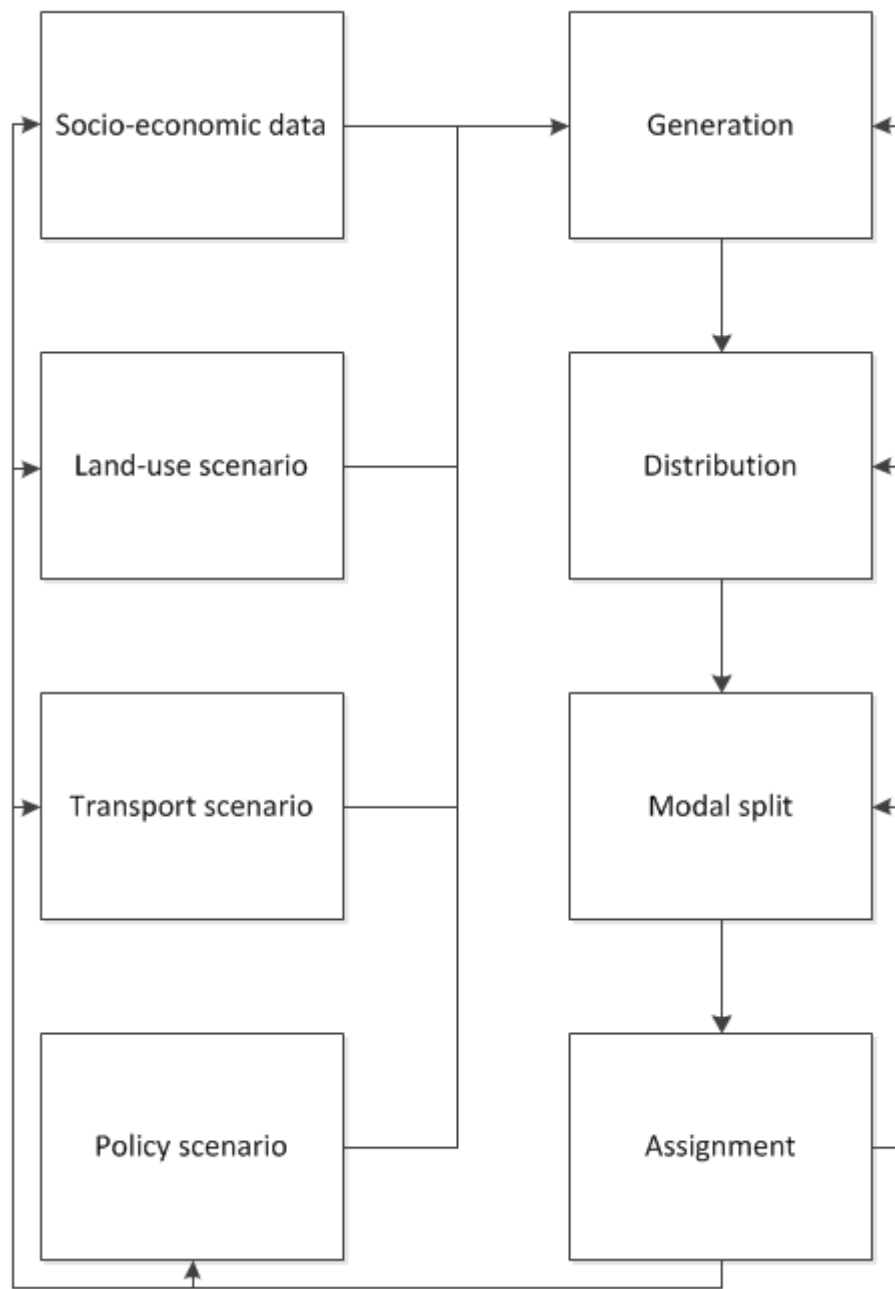


Figure 7: Illustration of the underlying concept in a traditional four-stage model.

engage in, but this first step results in the four-stage model being trip based rather than activity based. The generation is based on production and attraction of trips between different zones, and the magnitude of expected travel is thus determined by the amount of trip generators and attractors in each zone. This requires the network to be split into different zones, and intra-zonal trips are therefore often ignored. The dominant trip producer is typically the amount of households in a zone, while the dominant trip generator is typically the amount of workplaces in a zone.

The second of the four steps is the trip distribution, which aims at combining the production and attraction of trips for the purpose of obtaining a trip table (i.e. origin-destination matrix). This is

typically handled via a gravity model inspired by the Newton's theory of gravitation from physics. The basic premise is that there is a greater share of travel activity between zones in close proximity than between those that are far apart. However, due to the time-space compression in transport systems that is often caused by differences in geometry, technology, or politics, distance is typically replaced by a generalized travel cost. This cost is typically based on both physical distance as well as average travel time, but can in principle include other factors (e.g. comfort level). For car-based transport, road level-of-service is sometimes included in the generalized travel costs to reflect the attractiveness of a given route.

The third of the four steps is the modal split, which aims at splitting the trips on different forms of transport modes. Depending on the sophistication of the model this can be used to evaluate policy initiatives that aim at improving car-pooling options or non-motorized modes of transport, but outside larger urban areas it is not uncommon for the mode choice to be restricted to car use for the sake of simplicity. The last of the four steps is the assignment, which aims at splitting trips on specific routes in the system. So far the model has ignored the relationship between travel demand and network performance, but this can be handled in the assignment step through an iterative process (hence the arrows leading from this step in Figure 7). Rather than just assigning all travel activity to the fastest route, travel times for congested routes are updated for the next iteration to reflect the performance drop of the route due to the initial assignment. However, it is important to note that this typically only concerns response in the form of a different route or mode choice, while the trip generation is typically not affected. This results in a fixed origin-destination matrix (i.e. overall travel demand is unaffected, but in reality travellers might decide to work from home, move closer to work, travel outside peak-hours, or other responses that is not related to route choice).

The very brief outline of four-stage models given above is intended mainly to give readers a crude understanding of the dominant approach to demand forecasting, since the ex-post evaluation presented in chapter 6 is based on forecasts produced the trip-based approach in four stage models. Perhaps the most important point of this brief introduction is to highlight how this type of models simplifies relationships between endogenous variables through a sequential structure. Some of the limitations in this approach can be quite crucial to the validity of subsequent impact appraisals, which will be addressed more specifically in section 6.3. The problem is not so much that the models are inadequate at dealing with the full range of behavioural mechanisms that are relevant to travel demand, since any model is an intentional simplification of reality that necessarily have to omit certain mechanisms and use proxies for variables that are difficult to measure. Rather, the problem appears to be a common failure to fully acknowledge the implications of these simplifications in terms of what the model results can and cannot be used to construct valid inferences about. Even if four-stage models might be considered relatively transparent and intuitive from an engineering perspective (Rich 2010), the structural complexity of contemporary transport models can easily black-box the involved mechanisms and their relationships to non-experts (Hajer 1997; Henman 2002). This is especially the case since the outline of four-stage models given here is largely conceptual. Individual models are typically highly customized, and only a small group of people have the required technical competence and available resources to fully assess a given model. Policy makers thus depend more or less blindly on the results of consultant work and have to rely on peer review for validity and reliability checks.

2.6 Summary

The purpose of this chapter has been to introduce a set of concepts that are important to the topic of the present thesis. The stratified structure of reality has been described in line with critical realist ontology, and tendencies have been claimed to be a central concept in forming our scientific knowledge of reality. This laid the foundation for the introduction of models as necessary simplifications to explain complex systems, and led to a discussion regarding the level of accuracy to be expected in forecasts. Taxonomies for different types of uncertainty and accuracy were established, to distinguish between different dimensions of uncertainty and different forms of accuracy. A brief introduction of the dominant model type used for demand forecasting was included as in the end of the chapter, to illustrate the many simplifications necessary to arrive at a useful model. The main purpose of the arguments presented in the present chapter has been to highlight the inherent limitations for accurately predicting future states of social systems through model-based decision support.

3 Related works

If I have seen a little further it is by standing on the shoulders of Giants.

- Isaac Newton

In this chapter I will provide a brief overview of the previous works that have served as an inspiration to the present study. There is an ample body of literature available on best practices in forecasting and the impact that forecasts have on appraisal outcomes, and readers interested in such topics are advised to seek out some of the many textbooks that are available in these areas. I shall instead focus on the more practice oriented works that have tried to compare ex-ante forecasts with ex-post development. While forecasts play a prominent role in ex-ante appraisal for transport projects, this has not been reflected by a great deal of interest in post-auditing of forecasting practice, and the available studies are not always easily comparable due to a lack of standard approach to analysis. In this chapter I include what I consider to be the most relevant previous work for the present study, as well as references for additional studies that I will not cover in detail for various reasons that will be explained throughout this chapter.

3.1 Post-auditing of appraisal

As described in chapter 1, transport planning is a field studied by many different disciplines, and as a result the available post-auditing studies of transport projects focus on many different themes depending on the background of their respective authors. Audits by economists tend to focus on the wider economic impacts of the projects, engineers tend to focus on issues related to the physical construction and the environment in which it is placed, and managers tend to focus on process related aspects of project supervision and implementation. This makes it difficult to pool the results of different appraisals together to form a larger sample, as the items included in individual analyses are often not comparable. As an example, one study of the economic impacts from a project might only report the costs as total budget allocation and benefits as the first year rate of return. It would be difficult to compare the costs of this study directly to one that includes acquisition of land, construction, and operating expenses in its costs, just as it would be difficult to compare with one that doesn't report benefits as first year rate of return, but instead a cost-benefit ratio for the full lifespan of the project. A third study might not report the economic impacts at all, but instead focus on the impacts measured in absolute units such as the amount of fatal accidents or noise levels in decibel. In line with the demarcations listed in chapter 1, I shall primarily focus on studies of transport projects that have reported demand forecasts available at the time of decision to build and ex-post traffic volumes immediately after the project has opened.

Apart from a great diversity of authors in terms of the disciplines from which they approach these studies, it should be noted that the affiliation of the authors can also influence the scope and focus of auditing. This is particularly evident in the post-auditing of transport projects, where different studies offer different types of explanations for observed inaccuracy and bias in forecasting. Some audits are

performed by research institutes or individual researchers, while others are performed by governmental auditing agencies or internally in the responsible directorates (or private companies in the case of privately run infrastructure). Siemiatycki (2009) found that the former type of audits generally tend to focus on political or economic incentives for manipulating ex-ante appraisal in a way that fits better with the wishes of project proponents, while the latter type of audits generally tend to focus on design changes or delays in the construction process. This is likely due to a mix of differences in their mandates, professional background, access to data, and choice of temporal reference points. Auditors seem to be primarily concerned with the construction phase, while many scholars are more concerned with the planning and approval of projects prior to construction. It would also be naive to disregard potential economic or political interests from the authors in this relation, as internal audits obviously, *ceteris paribus*, must be expected to have less incentive to highlight lack of competences or deliberate fraud. Conversely, external audits have an incentive to fit the results with the perspectives of their clients, and scholars have an incentive to reach findings that match their own theoretical framework. Critics who claim that consultants have an incentive to promote a specific result are probably right, but the same argument could then be made for any other type of advisor, including the scholars who often form such criticism. Tal and Cohen-Blankshtein (2011) studied bias in forecasts by controlling for author affiliation, institute, and publication type, but found that none of these characteristics had any influence on forecasting bias. Researchers were found to be just as likely to produce biased forecasts as consultants and government agencies, and bias is just as likely to be present in scientific publications as in appraisal reports. Tichy (2004) also note that experts generally tend to bias towards optimistic forecasts, although the magnitude of bias is typically weaker in academia than in business environments. There is no easy way to handle these issues, and auditing the auditing reports themselves is often an almost impossible task due to limited data availability. It would also risk distancing appraisals too far from their actual units of analysis, since such investigations would require many uncertain assumptions to be made. At any rate, it is far beyond the scope of the present thesis. Nonetheless, it is an important acknowledgement to keep in mind when assessing the possible explanations offered for inaccuracy and bias in forecasting, just as it becomes a key determinant for the solutions offered to deal with these problems.

In relation to forecasting inaccuracy, there seems to be broad agreement in academia, politics, and the media that cost overruns and benefit shortfalls for public works projects are not uncommon. In fact, many probably consider it to be a rule rather than an exception when they hear yet another story of a public works project that didn't meet expectations, whether it is in relation to the anticipated benefits, initial budget, or deadline for completion. However, large-scale studies of actual bias in a representative sample are fairly rare phenomena, and transport projects are certainly no exception. Many of the previous studies include either relatively small samples that are often difficult to generalize from, or are themselves aggregated results from other studies with small samples. For smaller samples there is a clear risk of bias from an availability heuristic, where projects are selected due to the very nature of them having inaccurate forecasts. These projects are often easier to recall due to our memory being biased towards vivid or unusual phenomena, and probably receive a larger amount of attention than their actual population share would merit. Skaburski and Teitz (2003) and Siemiatycki (2009) argue that this is a likely cause of observed forecasting inaccuracy in planning. This

bias is well documented in psychology, and Tversky and Kahneman (1973) found that when people are presented with a list of items, they have a much higher chance of remembering items that stand out than is the case for more mundane items. When asked to recall how big a share of a total population these more recognizable items make up, people typically reveal an inflated perception of their actual representativeness. Since transport infrastructure projects are usually large public works projects, which might be heavily contested in some parts of the public, cost overruns and benefit shortfalls will naturally attract a lot of attention from various interest groups. As a consequence, these projects are likely to gather significant media attention, and will probably be subject to much more critical scrutiny than projects that do not experience these problems (or at least where performance criteria are less obvious). I will therefore claim that transport infrastructure projects with problematic forecasts are likely to be overrepresented in the general media picture, and this is probably also true for some of the post-auditing studies that have been done in this area. In fact, some of these studies are probably undertaken as a direct result of problematic forecasting in the projects included in their samples, meaning that data availability could also be higher for projects with inaccurate forecasts.

The availability heuristic can create an inherent bias in the selection of items, which is of course crucial to studies of systematic biases in the results of forecasts. If the sample is constructed from systematically biased selection criteria, then the results will of course also show a tendency of systematic bias, even if this is not true for the general population. This need not be a problem in the individual auditing study, as the entire purpose of it could be to investigate the causes of these problems in the specific projects under scrutiny. However, it would be dangerous to aggregate the results of many smaller studies, since this could result in systematically biased results due to the aforementioned selection criteria, and thus not be a true reflection of the magnitude of any inherent problems in the project population as a whole. When pooling results from different samples one must thus be aware that the individual projects for these samples must be randomly selected if we are to expect the aggregate sample to be so as well. As smaller samples are often not random by design (or intended to be so), this is unlikely to be the case if it is not specifically corrected for when pooling results.

Another important issue when comparing the different post-auditing results is the different spatial and temporal settings for the projects involved. Different audits might be based upon samples of projects from different periods in time or different geographical locations, which raise a number of questions. Were the projects constructed in periods of economic growth or recession? Did the political climate favour or oppose the projects? Are they located primarily in urban or rural settings? Are they new projects altogether or expansions of existing infrastructure? These are but a few examples of possible factors that could influence the results, and like many of the other issues that have been raised so far, they require quite detailed knowledge of the samples, as well as access to data that is either not available or quite resource-demanding to come by. However, inability to control for such factors due to limited access to data does not mean we can simply ignore these issues completely, and I shall therefore devote some time to them when going through some of the related works

Before introducing some of the previous work on post-auditing of transport projects, I should note that the studies selected here are by no means a comprehensive list. This is primarily due to the following two reasons. First, while I have claimed that this is often a neglected field of study, there are too many such studies to cover in detail here, and many of them have very small samples or confine themselves to studying a single project. With the exception of Kain (1990), I have intentionally excluded post-audits of single projects, since it would be unrewardingly resource-demanding to include a large number of such studies here as well as inflate the risk of selection bias. The intention is not to give a comprehensive overview of all past studies or their results, but to illustrate the differences in scope and analysis that these studies represent. Second, many of the post-audits cover very different topics than those highlighted in chapter 1. While it might be fruitful to discuss these differences and their implications, I have chosen to exclude a detailed presentation of them in the present chapter, and shall instead refer to them where they are relevant to the arguments presented.

The studies presented in section 3.2 are thus included either due to their alignment with the purpose of the present study to allow a comparison with the results presented in chapter 6 or to illustrate important methodological concerns that need to be addressed. It can of course be contended whether I have left out important studies or included some that are of little relevance, but any non-exhaustive list would be prone to such critique. The studies I have included are the ones I find to be the most commonly cited in the academic literature produced during the new millennium. The reason for this temporal cut-off is that there has been a growing attention on performance evaluations since the 1990s, which culminated with the Ph.D. thesis of Mette Holm (2000) and the subsequent popularity of publications based on the material presented herein. Most notable are probably the work of Bent Flyvbjerg and his colleagues (Flyvbjerg, Bruzelius, and Rothengatter 2003; Flyvbjerg, Holm, and Buhl 2002, 2005, 2006; Flyvbjerg 2007), who was the supervisor of Mette Holm during her Ph.D. studies. These works continue to be some of the most cited pieces of literature on both cost and traffic forecasts for transport infrastructure projects, and have also been a great source of inspiration for the present study. For these reasons the work by Flyvbjerg et al. will also be given more attention than the other studies presented in section 3.2. It seems to have created a surge of interest in post auditing, and I have thus chosen to focus on larger studies conducted after this work, as well as any prior analyses that are continually referenced in these more recent studies.

As traffic forecasts are the main focus of the present thesis, I shall omit a comprehensive description of the literature that deals exclusively with the accuracy of cost forecasts for transport projects. However, it bears repeating that traffic forecasts and construction costs are the two most crucial aspects of CBAs in the transport sector, and that a systematic bias of cost underestimation seems to be the standard for projects in many different sectors. In fact, large infrastructure projects appear to be especially problematic in this regard, where cost overruns are typically more severe than for other public works projects (RR 2009). A great deal of research has been directed at this topic, and I refer to Siemiatycki (2009) for a well-structured overview of the most influential studies in this field, as well as a summary of their conclusions. Furthermore, his selection criteria for cost studies appear to be similar to those I employ for traffic studies, and most of the studies in his comparison include large samples of projects. The general, and perhaps unsurprising, finding of Siemiatycki's review is that cost overruns are very common indeed (average actual costs are larger than average estimated

costs in all studies), and that the degree of bias is quite significant (50% cost overruns are common for non-routine projects). Overall, these findings confirm a general attitude amongst planning stakeholders of cost overruns being the rule rather than the exception for larger public works projects. The reason I label the results as unsurprising is that such findings are neither new nor limited to transport projects (see e.g. P. Hall 1980; Kharbanda and Stallworthy 1983).

3.2 Studies of forecasting accuracy

In this section I will introduce what I consider to be the key studies on the accuracy of demand forecasting. For each study I shall attempt to outline the spatial and temporal boundaries for the projects included, as well as report the general imprecision and bias in the form of the distribution of inaccuracy for individual samples. However, as there is not a standard convention for how to assess the performance of demand forecasts, it might not be possible to report comparable items for all studies. I shall try to bring attention to the most important methodological differences between the studies presented in the present chapter, and where possible I have tried to convert the most important variables into a comparable format.

An overview of the studies can be seen in Table 1. The sample size listed here is for the entire pool of projects included in the individual studies, although for some items data might only be available for a subset of the sample. The sample size indicated for analyses on individual items from these studies in their respective subsections might therefore differ from the sample size in Table 1. There are of course many other available comparisons of expected and actual traffic, but many of these do not have sufficient documentation for the necessary items covered here, or their small sample sizes are too small to be of interest. Brinkman (2003) offers a good discussion on many of these studies as well as some of their main limitations, although he appears somewhat inconsistent in his evaluation criteria for the validity of individual studies (e.g. criticizing some studies for their choice of methodology but uncritically accepting findings from other studies with similar methodology). It should be noted here that one study in the sample for the present study (Mackinder and Evans 1981) measures inaccuracy in a very different way than the one I described in chapter 2, and instead employs the following measure¹²:

$$I_s = \frac{B + (P - B) \left(\frac{10}{T_P - T_B} \right) - B + (O - B) \left(\frac{10}{T_O - T_B} \right)}{B + (O - B) \left(\frac{10}{T_O - T_B} \right)} \times 100$$

It would be desirable to make such an extended model (I_s) of inaccuracy for all studies, as this takes into account the estimated change over a span of ten years¹³, and would thus account for differences in design year and opening year (I shall elaborate on this latter problem in 4). However, this would require data items that are seldom available for studies of this kind, and doing so would likely cause the interpretation of results to be less intuitive to readers, whereas the simpler model employed in

¹² B = base value, P: predicted value, O: observed value, T_x: year of measure for variable X

¹³ In Mackinder and Evans' version this span is calculated by a linear model, but the format could easily be converted to an exponential model. This would typically inflate the inaccuracy measure slightly, but likely result in a better representation of the typical assumption for development trends in transport infrastructure appraisal.

Author(s)	Period	Area	Sample ¹⁴	Mean	Std. dev.
Mackinder and Evans (1981)	1970s	United Kingdom	Road: 44	-7% ¹⁵	N/A
National Audit Office (1988)	1980s	United Kingdom	Road: 161	+8%	43
Pickrell (1990)	1980s	United States	Rail: 10	-65% ¹⁶	17
Fouracre et al. (1990)	1980s	Developing countries	Rail: 13	-44%	26
Kain (1990)	1980s	Dallas	Rail: 1	N/A	N/A
Flyvbjerg et al. (2005)	1970-90s	Global	Road: 183 Rail: 27	+10% -40%	44 52
Bain (2009)	N/A	Global	Toll: 104	-23%	26
Parthasarathi and Levinson (2010)	1960-2000s	Minnesota	Road: 108	+6%	41
Welde and Odeck (2011)	2000s	Norway	Toll: 25 Road: 25	-3% +19%	22 21

Table 1: Overview of past studies of traffic forecasts inaccuracy.

the present study (I_p , see section 2.4) is very easy to interpret even for people who are not particularly math savvy. I have therefore converted all inaccuracy estimates to the form I_p where the available data has made this possible. The values presented for Mackinder and Evans (1981) in Table 1 are converted to the I_p format, but under the individual description of the findings in this study such conversion has not been possible, and the I_s form will have to be used instead.

A few things should be evident from the results in Table 1. First, there seem to be a quite clear tendency of overestimation of traffic for rail and toll projects and underestimation of traffic for road projects, as indicated by the mean value of the project samples' inaccuracy. This means that, on average, there are fewer users of rail and toll systems that anticipated at the time of decision to build them, while the reverse is true for road projects. Second, the sample size for rail studies is typically

¹⁴ The three sample types refer to rail projects (light, heavy, metro, etc.), road projects (highway, bridge, tunnel, etc.) and toll projects (same as road projects but with direct user charges).

¹⁵ This value is for screen lines rather than individual stretches due to lack of data for the latter.

¹⁶ If total transit trips rather than total rail passengers is used as a reference the value becomes -41%.

much smaller than for road and toll studies, which is probably a reflection of the population size for the different types of infrastructure projects. In addition to this, due to the often complex nature of rail systems, with many different owners and service providers, reliable data sources are often much easier to access for road projects than for rail projects, since there is typically a single federal agency that administers most of this information for roads. Third, over half of the studies limit themselves to a specific geographical context, typically in the form of a single country. The exceptions are studies by or involving private companies (Bain 2009; Fouracre, Allport, and Thomson 1990) and aggregate samples of results from several previous studies (Flyvbjerg, Holm, and Buhl 2005). Fourth, there seem to be a fairly high inaccuracy reported in studies of both new and old projects, as indicated by the standard deviation of the project sample's inaccuracy (std. dev.). This implies that forecasting accuracy have not improved a great deal over the last four decades, and that many transport infrastructure projects have been decided on the basis of highly uncertain decision support.

It is of course difficult to generalize based on a fairly small sample of studies, of which many employ very different methods for assessing the inaccuracy of traffic forecasts. However, the immediate observations listed above at least indicate that traffic forecasts are often highly uncertain, and that there seem to be a considerable amount of bias in demand forecasts for the rail sector. In the remainder of the present chapter I shall elaborate a bit on the methodological differences between the individual studies presented in Table 1, as well as discuss a set of considerations for the present study based on the findings herein. As mentioned earlier, there seem to be no standard convention for a methodology of ex-post appraisal of traffic forecasts. This obviously makes a comprehensive cross-study comparison problematic, since the underlying concepts for any analytic categories in such a framework will have to be broad enough to encompass studies with very different appraisal scopes, data items, and project types. The categories I have chosen are thus primarily a set of concepts that have emerged through a somewhat semi-structured literature review, and are perhaps partly inspired by an informal grounded theory approach¹⁷.

3.2.1 Mackinder and Evans

Mackinder and Evans (1981) did a comprehensive study of 44 UK road projects completed in the 1970s. A couple of earlier studies on forecasting accuracy are available, but have been omitted as these are mostly referenced in Mackinder and Evans's work. Furthermore, the 1981 study also better represents the geographical focus area of the present study. In addition to this, their work includes a larger amount of projects and better sources of data than many of the previous studies, as they had direct access to the necessary data from local transport planning authorities. This allowed a thorough analysis of the forecasting performance of twelve individual items, including both highway and transit traffic.

The results showed a clear tendency of overestimated forecasts for 12 items analysed in the study. Highway traffic was overestimated by as much as 30% on average, while transit trips were overestimated by 35% on average. Interestingly, the median values (21 % and 35% respectively)

¹⁷ While I consider grounded theory to be faced with quite significant epistemological challenges, the general premise of an explorative approach still seems useful in areas where little established theory exists. I do not consider grounded theory a well-established methodological framework, but as a set of guiding principles I believe it suffices for the sort of analysis that is presented in this section.

differed much more than the mean values, indicating that quite severe forecasting inaccuracies for certain road projects distorted the mean. Highway screen line trips were much less pronounced, with only 13% overestimation on average (median of 8%). However, these screen lines were only available for a few studies, and for this subset the general highway traffic seems to be less overestimated as well. Part of the discrepancy in the accuracy of these items thus seems to be stemming from the use of different samples, although no specific project identification is provided for these groups by the authors to verify this assumption.

Mackinder and Evans attribute much of the forecasting inaccuracy for traffic volumes to optimistic national growth forecasts for a range of input variables (see Figure 8). The errors in car ownership and income levels are examples of such variables, which were both overestimated by roughly 20%. Land use items such as population, employment, and household size were generally less inaccurate, but all of them were overestimated as well. A rerun of the models with these new input values indicated that inaccurate input variables could explain most of the inaccuracy in highway traffic forecasts. This was not the case for transit trips, where the authors point to specification error instead. Most of the models use a post-distribution modal split, which estimates the total trip volume before assigning modes for individual trips. However, while highway trips did indeed increase, this was evened out by a reduced transit usage. Even with correct input variables this model specification would have resulted in overestimated transit trip forecasts.

Perhaps the most interesting observation from this study is the fact that outturn traffic for road projects is below the expected values on average, whereas all other large studies of forecasting inaccuracy for road projects show a reverse trend of general underestimation. One possible explanation for this seems quite obvious. Many of the projects in this study were decided in the 1950-60s, but were completed in the last half of the 1970s. The decision support have thus been prepared in a period of relatively strong economic growth, while the opening years lie between two oil crises in a period of global recession and stagflation. All else equal, this should result in traffic being lower than forecast. Figure 8 provides some support for this explanation, since household income, employment rates, and cars per capita estimates were overestimated, all of which are very likely consequences of a recession. The marginal costs of travel were also higher than expected (probably partly due to the 1973 and 1979 oil crises), although comparisons for these items are scarce reported by Mackinder and Evans (1981). However, the authors note that the majority of models at the time assumed trip rates to be independent of travel costs, in which case there would not be much point in controlling directly for such a variable anyway. In spite of being excluded from the model structure, it is important to keep in mind that increased travel costs could still be a very important explanatory factor for the observed traffic forecast inaccuracies, since such a specification error more or less equals an assumption of status quo travel cost, which turned out not to be the case.

3.2.2 National Auditing Office

Partly inspired by the findings in Mackinder and Evans, the UK National Auditing Office (NAO 1988) did a study of 161 road projects completed in the period 1980-88. The projects in the study make up 68% of the total projects completed in England and Wales during this period, while projects from Scotland and Northern Ireland are not included. They were primarily concerned with comparisons of

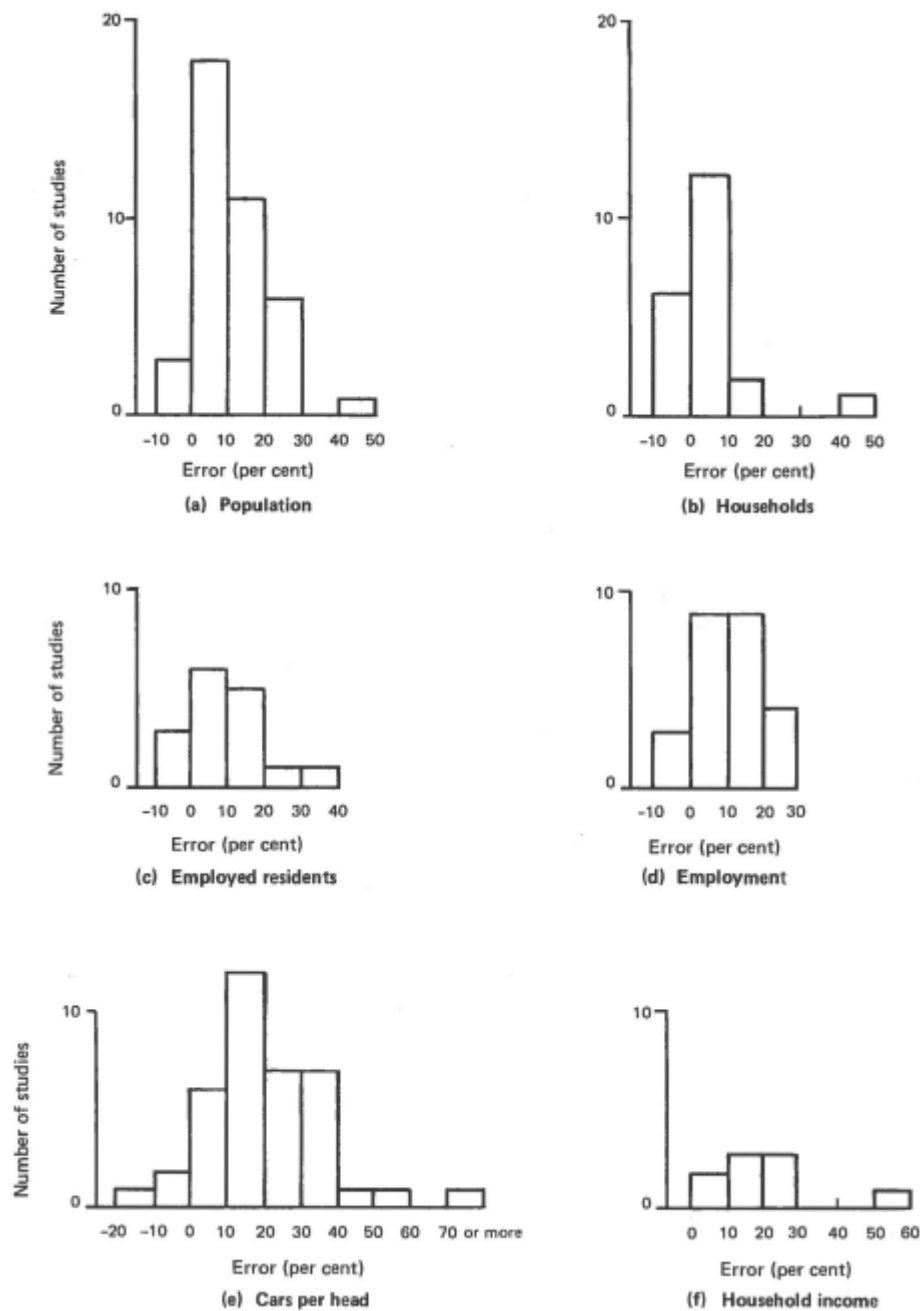


Figure 8: Inaccuracy distributions of the key input variables in UK road projects. Note that inaccuracy is measured as I_R instead of I_p for this figure, meaning that positive values represent projects with overestimated forecasts. Reprint from Mackinder and Evans (1981).

forecast and estimated traffic flows in the design year, and thus square well with the approach adopted in the present study. However, out of the 161 projects only 124 have data available for a direct comparison with the rest of the studies presented in the present chapter, since actual figures are not reported directly and have to be assessed through graphs.

Going through the available data reveals an average inaccuracy of 8%, meaning that traffic flows were generally larger than expected. The spread of inaccuracy is quite large though, and individual

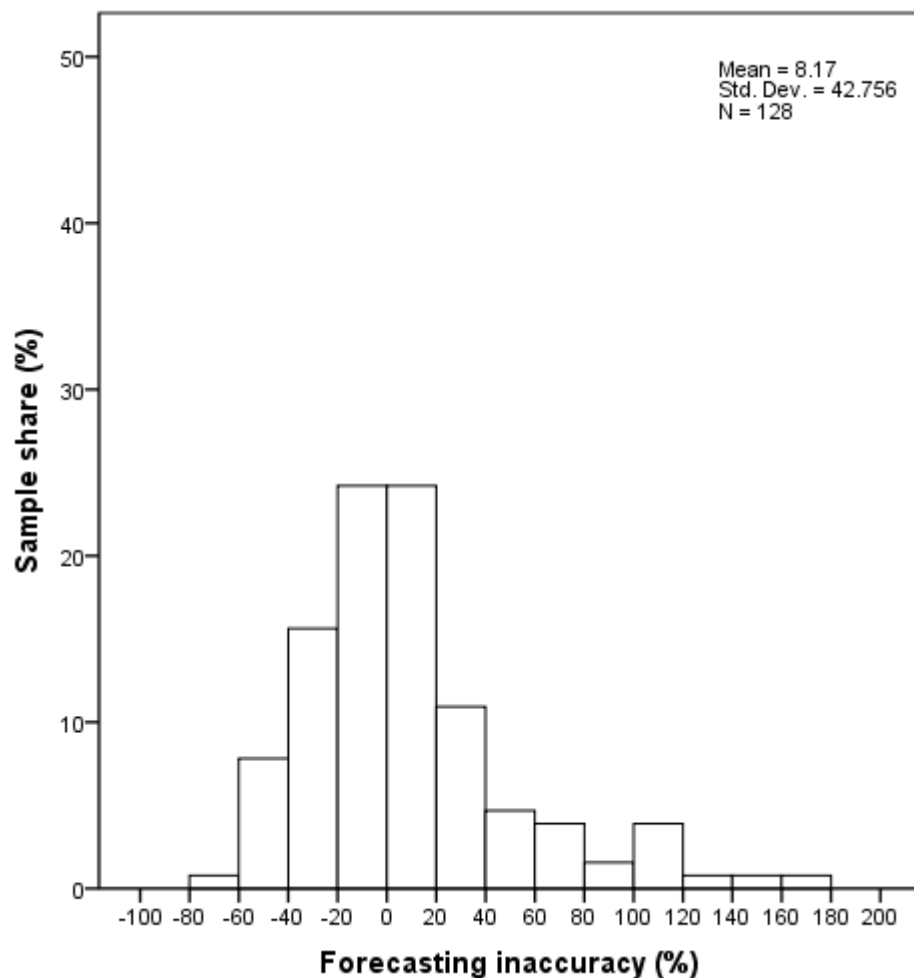


Figure 9: Inaccuracy of traffic forecasts for UK motorways. Data source: NAO (1988).

projects fall in the range of -67% to +161% (see Figure 9). Moreover, there seem to be a trend of Northern and Western regions primarily overestimating traffic, while Southern and Eastern regions primarily underestimate traffic. The authors attribute this marked regional variation in accuracy with similar variations in GDP growth, as the traffic forecasts were based on the expected growth rate across all regions. The regions experiencing more traffic than expected are also the ones that have experienced GDP growth above the expected national average. However, since very little actual data is presented in the study, it is difficult to verify the impact that these erroneous assumptions might have had on the expected traffic flows.

The intra-regional span of uncertainty on individual road stretches is certainly far too large for mere differences in GDP growth to be the sole cause of inaccuracy. The NAO lists a set of contributors to the observed deviations between predicted and estimated traffic volumes, of which three are unrelated to GDP growth. First, traffic diversion to new motorways and bypasses in rural areas have been overestimated, since a lot of drivers continue to use the old road due to a lack of facilities along the new roads. Second, the effects of induced and redistributed traffic have not been given sufficient attention, and fixed matrices have been standard in many appraisals. Third, unexpected

developments have often affected project outcomes, and especially changes in the planned development of land use after the preparation of forecasts are mentioned as a cause of this.

The results in NAO (1988) are quite important due to the large sample size of the study, which covers the majority of all projects completed in the target area during the 1980s. Furthermore, in contrast to Mackinder and Evans (1981), but in line with subsequent studies, the authors find a positive rather than a negative average inaccuracy for road projects. It is tempting to look for explanations in the economic trends of around the time of decision-making and opening years respectively, and conclude that projects were decided in the aforementioned stagflation period and completed under more favourable economic growth conditions. Part of the explanation also seems to involve the exclusion of induced demand as a result of faster travel times. However, while both of these are plausible explanations for the general underestimation of traffic, they do not really address the large spread of inaccuracy that is also observed in this study. Unlike the data from Mackinder and Evans (1981), the NAO (1988) study allows the calculation of a standard deviation (see Figure 9). Considering the relatively large sample the deviation must be considered to be rather high, and indicates that there are some major uncertainties involved in the prediction of future travel demand. It seems clear that the transport models have either been specified based on a flawed understanding of travel behaviour or fed with incorrect assumptions regarding the development of key input variables. According to the authors, both of these statements are true for many of the projects in the sample, and judging from the simplified model structures combined with poor estimates of economic growth, national traffic growth, and land use development, these seem like quite plausible explanations.

3.2.3 Pickrell

Pickrell (1990) did a comprehensive study of 10 urban rail projects in the US¹⁸ completed in the 1980s, which continues to be referenced in many contemporary studies on forecasting accuracy in the rail sector. It has been quite influential in the US, where it has been a staple item in the repertoire of transit critics due to the severe overestimation of ridership that can be observed in the results. Most of the projects did not even achieve half of the forecast ridership in the opening year (see Figure 10), just as most of them were also far over budget. None of the projects managed to attract ridership equal to or above the levels indicated in the forecasts.

Just like Mackinder and Evans, Pickrell also examined a range of input variables used in the forecasts. The results generally indicated that important input variables were often given too optimistic values. In general, population in the service area was overestimated, headway was underestimated (especially for feeder services), and operating costs of cars were overestimated. The latter of these is probably a result of a general concern with escalating oil prices in the 1970s, which lead planners to severely overestimate operating costs of car travel by extrapolating this short term trend. However, for a couple of projects parking prices were also underestimated, which has partially offset the bias from overestimated operating costs.

¹⁸ Out of these ten projects four can be classified as conventional rail, four as light rail, and the last two as downtown people mover systems.

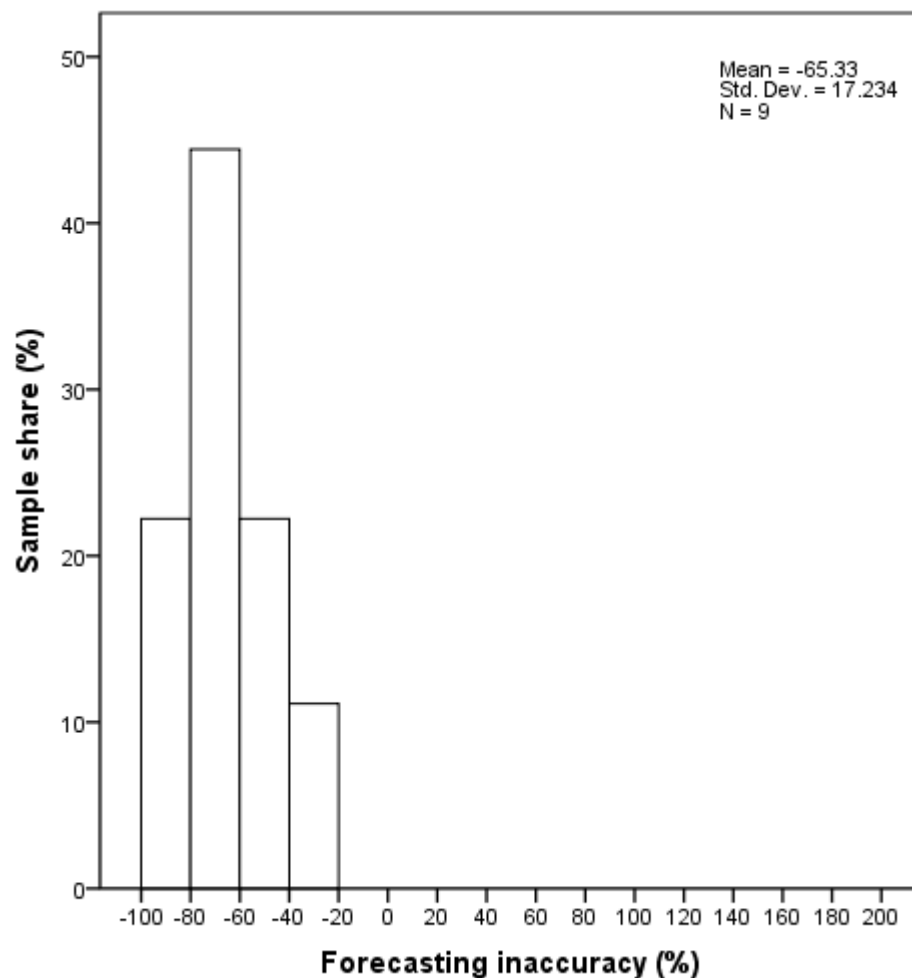


Figure 10: Inaccuracy of traffic forecasts for US rail projects. Data source: Pickrell (1990).

Pickrell tried to assess the impact of faulty assumptions for input variables on the ridership forecasts, and came to the conclusion that inaccurate input variables could not have accounted for more than half of the observed errors at best. However, as he did not have access to the actual data used in the transport models, he chose to use published elasticity values from two conference papers for many of the variables, while others were just assumed to be directly proportional (i.e. elasticity values of 1.0 in respect to ridership). While an acceptable compromise under such conditions, the reappraisal method used by Pickrell for gauging the impact of inaccurate input variables is problematic in the sense that we do not actually know if the models used to produce the forecasts would behave similarly, as they might have used very different elasticities. These items are rarely reported explicitly in appraisal reports, which make it a nearly impossible task to control for such uncertainties. However, that does not lessen their potential influence. For example, if elasticities of parking costs in the original forecasts were in line with those reported in TRACE (1999) rather than the conference papers referred to by Pickrell, this would mean that the forecasting errors for this variable could explain a larger share of the ridership overestimates than Pickrell's reappraisal concludes. Without any knowledge of the actual values used in the transport models this type of reappraisal is mostly an educated guess, as elasticity values can vary greatly between different settings. As an example, APTA (1991) reports transit fare elasticity values in the range of 0.12 to 0.86 between different cities in the

US. It is very difficult to evaluate the impact of variables with such large spans when we do not know the exact value used in the forecast.

Another example is bus headway, which has a forecast span of 2-40 min for one project in Pickrell (1990), while the actual headway turned out to be 5-15 min. I have tried to reconstruct Pickrell's reappraisal using the elasticities he reports, and if the minimum headway values are used the appraisal shows inaccuracy in this variable to account for all of the ridership inaccuracy. If we compare by the mean values instead, the appraisal shows that headway inaccuracy alone would have caused even more inflated forecasts. The minimum headway is likely an expression of rush hour conditions, which in my opinion makes it the most suitable candidate for the kind of reappraisal undertaken here, as most of the passenger flow takes place during rush hour. However, Pickrell seems to have chosen the mean value as a basis for comparison instead, which would lead him to conclude that this accuracy could not explain the inaccurate ridership forecasts. I do not agree with this interpretation if that is really the case, but the point of presenting both arguments here is not to judge which of them is more correct. Rather, they should serve to illustrate the caution we must take when evaluating performance in light of scarce data access, since it is possible to arrive at conflicting interpretations depending on the assumption we make regarding things we have no certain knowledge of. This is true for project appraisal as well as appraisal of project appraisal.

On top of all the uncertainties regarding elasticity values, we also lack knowledge of the model structures, as well as information about how the transit networks have been designed and coded within them. As Pickrell mentions himself, errors from such sources are extremely difficult to detect, but might have profound influence on the forecasts. I shall not go into detail with these issues here, as it is possible to employ a similar argumentation as that presented above for why this would yield little or no insight. For now, let us just note that it is practically impossible to rule out the possibility of model specifications to have been the cause of inaccuracy, since we do not have access to this type of information.

The concerns raised here is not an attempt at belittling Pickrell's study or the results presented in it, as I find it to be a very thorough analysis of the selected cases. However, there is reason to pay great attention to these conditionals when we discuss the implications of the findings. The report has served as a bit of a bible for the anti-transit lobby in the US¹⁹, particularly because Pickrell hinted at the possibility of "*misrepresentation of numerical outputs during the planning process*" as a source of forecasting inaccuracy²⁰ (Pickrell 1990, 29–30). Meanwhile, the nuances displayed in Pickrell's original assessment sadly seems to have vanished from the debate it sparked, which have predominantly focused on the speculations of deliberate fraud by project proponents. This is certainly a possible explanation, but mere motive for a crime cannot be taken for conclusive evidence, and Pickrell (1990) did not claim to have empirical evidence to give strong support for such accusations.

¹⁹ For a more elaborate discussion of this issue, see Weyrich and Lind (2001).

²⁰ Obviously this was not the only possible explanation offered in the report, but it certainly seems to be the one that has gained the most widespread attention.

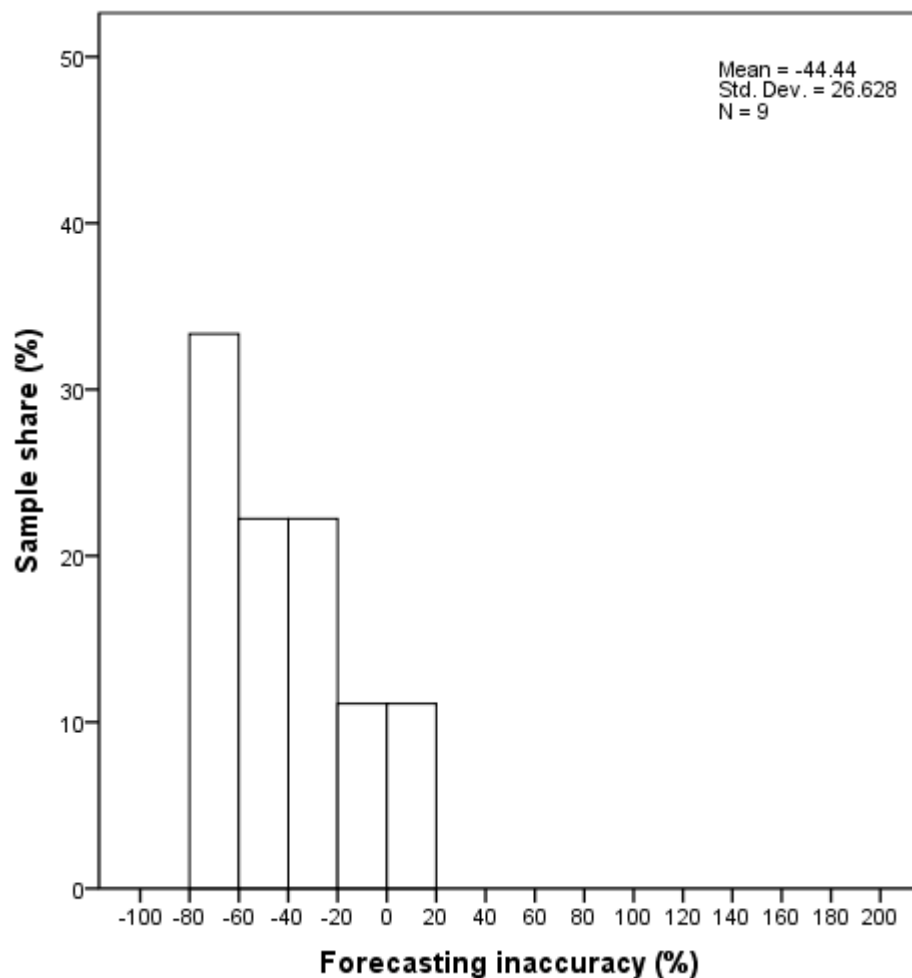


Figure 11: Inaccuracy in traffic forecasts for rail projects in the developing world. Data source: Fouracre et al. (1990).

3.2.4 Fouracre, Allport and Thomson

Fouracre et al. (1990) compared forecast and actual traffic for 13 metro projects in developing countries. It was intended to scope the circumstances under which a metro system would be a feasible investment for larger cities in the developing world. The approach appears very similar to that of Pickrell, but with much less detail given to data items and with no clear indication of data sources. This is not unique however, and Pickrell's work is perhaps the odd study that has documented data sources extremely meticulously.

The findings by Fouracre et al. are quite similar to that of Pickrell's study. Capital costs are generally underestimated, and traffic forecasts are generally overestimated (see Figure 11). Since the scope of the study was not primarily orientated towards comparisons of forecast and actual ridership, there are few options of controlling the values for possible influential factors. This makes it difficult to do much in-depth analysis of the possible sources of errors related to input variables, apart from what is listed in the study itself. The authors cite overly optimistic expectations in the planning process as the main source, as well as poor integration and choice of alignment.

The authors also examined the reasons for building metros in the case study cities, to gauge whether the projects might have provided the expected benefits in spite of falling short of forecasted ridership values. The two most prominent reasons for implementing metros were to raise the quality of public transport and to relieve traffic congestion, with the former being particularly strong in Latin America and the latter in the Far East. Financial and economic viability ranks third and fourth among the listed reasons, and the metros were expected to attract people and jobs to the areas in which they were to be placed, while reducing the amount of accidents and lowering emissions levels.

If we look beyond the traffic forecasts, it seems that some of the projects managed to fulfil the aforementioned goals quite well. The quality of public transport was certainly improved in most cases. As an example, the average metro journey of 15 minutes substituted an average bus ride of 40 minutes for one project, with superior levels of service for comfort and security. The authors note that while the metros can experience overcrowding during peak periods, the bus systems they replaced would be far below the necessary capacity to transport the volumes of traffic that the metro now serves. Meanwhile, none of the cities experienced any long-term reduction in congestion on the road network. Some short-term improvements were observed, but these quickly vanished. The authors attribute this to the effect of induced traffic.

Perhaps the most interesting results of the study are those presented in the economic analyses, which show that, in spite of cost overruns and ridership shortfalls, many of the projects have actually come out with a positive net present value and decent internal rates of return. Most of the projects that experience negative net present values do so because of low income levels, since most of the benefits stem from time savings and are thus proportional to the income levels. This is not the case for costs (at least not to the same degree). The authors conclude that few metros in developing countries have the possibility of being financially viable, but that there are still economic gains to be had from them under the right conditions²¹. In addition to this, many of the lines that are judged to be economically inefficient could have proved much more feasible with better planning in the implementation phase. This was typically manifested through poor integration with bus lines, neglecting the importance of fare structures, and poor choice of alignment. Both the first and latter of these items are illustrated well in the case of the Porto Alegre Metro, where planners tried to exploit an old railway line running parallel to a major highway. Much of the ridership was expected to come from new urban development that never materialized, resulting in a very poor catchment area. In addition to this, the metro had to compete with the existing bus lines on the highway, but as the metro terminates relatively far from the city centre there is little competitive advantage when compared to bus alternatives that proceed downtown. While the choice of alignment on an old rail line might have reduced construction costs, this compromise resulted in such a poor utility value of the metro that it was probably bound to perform poorly both in financial and economic terms.

²¹ Some of the conditions put forward in this study should probably be considered specific for developing countries. There are several reasons for this, but perhaps the most important one is the lower income levels that characterize developing countries. Since the economic benefits are proportional to income levels, recommendations for certain conditions (e.g. the desired ridership potential) might differ greatly from those that would apply to more developed countries.

3.2.5 Kain

Kain (1990) did a case study of a light rail project proposed by Dallas Area Rapid Transit (DART) and North Central Texas Council of Governments (NCTCOG) in the 1980s. This study is not fully comparable to the other studies presented in this chapter, due to it being only a single case rather than a pool of projects. Moreover, it is a case of a project that was in fact not built at the time of the study's publication, and therefore includes no comparison of forecast and actual ridership. However, I have chosen to present it here as it is an oft-used reference by scholars studying the causes of overestimated ridership forecast.

The project under scrutiny in Kain's study was a proposed light rail system in Dallas, and while there were no actual ridership volumes to compare with, he did a fine job of describing the development of the key forecasts that were presented by DART/NCTCOG. The first of these were the central business district (CBD) employment forecasts, which averaged somewhere around 200,000 jobs in most of the analysis put forward by DART/NCTCOG. Meanwhile, actual CDB employment estimates from counts had never exceeded 130,000 in the period 1958-88, with a steady decline through the 1980s bringing the number closer to 90,000 around the time of Pickrell's analysis. Given the complete lack of CBD employment growth over the past 30 years, Kain found it highly suspicious that some of the estimates expected CBD employment growth of more than 100% in the years to come with no explanation for why such a development was to be expected.

Kain also found that DART/NCTCOG had been hiding forecasts that did not support their proposed rail project, such as forecasts for a much cheaper bus solution, which indicated a very similar ridership potential as that of the proposed rail system. In addition to this, DART/NCTCOG released forecasts for three different scenarios to give the impression of conducting sensitivity analyses, while even the most conservative forecast was in fact highly optimistic in its growth assumptions. The continued bias in all forecasts released by DART/NCTCOG, as well as the active attempts to conceal unfavourable forecasts, led Kain to conclude that Dallas voters and policy makers could not trust these organizations to provide honest assessments. Instead, he claimed that DART/NCTCOG deliberately tried to mask their own preferred rail solution as the only feasible solution, despite several analyses indicating otherwise.

It is clear from the study that Kain has a great deal of intimate knowledge of the case he investigates, and has been monitoring the development of it as well as getting personally involved along the way. This allows him to present detailed descriptions of the process with references to both public and internal documents produced since the project was first proposed. While he presents no smoking gun, he probably gets as close as possible to conclusive evidence that forecasts have intentionally been based on overly optimistic assumptions, and that unfavourable assessments have been subdued where possible. It seems impossible for planners to innocently make this many leaps of faith that all point in one direction, and Kain is therefore convincing in his claim that the figures have been constructed with the intention of swaying public and political support in favour of the project. To me, the most striking feature of the Kain's study is probably the amount of intimate case specific knowledge that has been presented, while we are still left without conclusive evidence. This makes the study a clear example of the difficulties associated with proving intentional deceit as a cause of forecasting inaccuracy. Few people would have access to this kind of detailed knowledge in a case

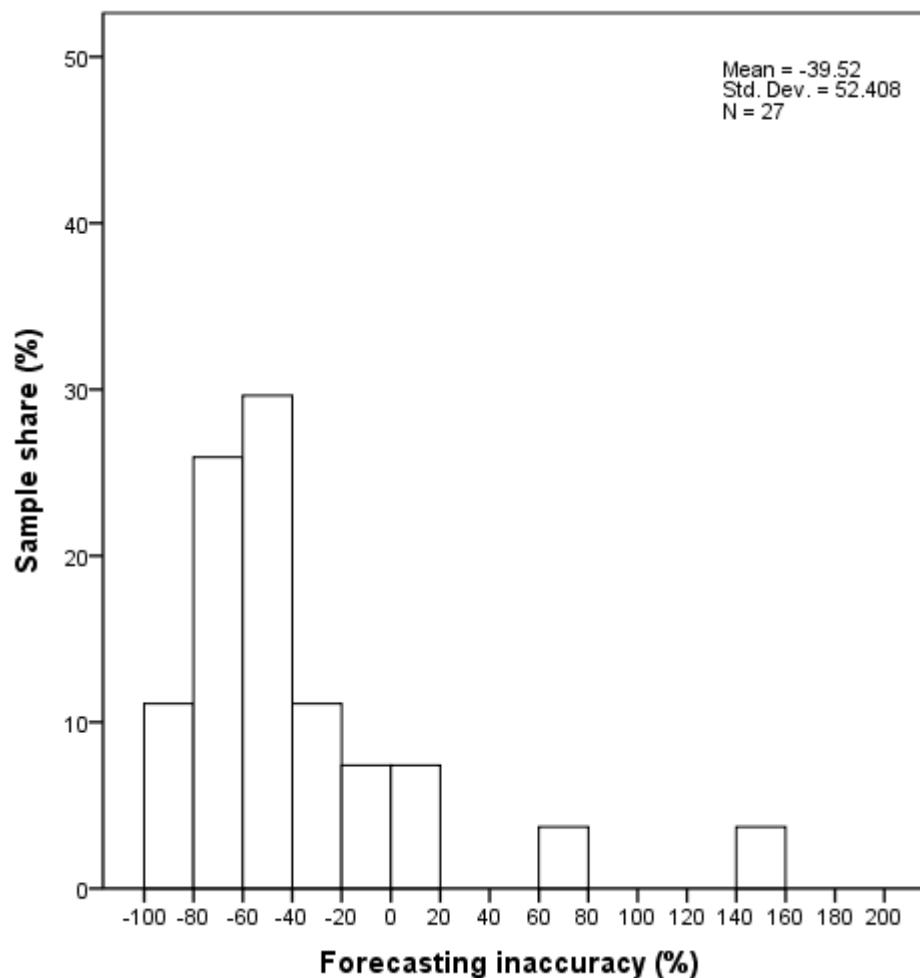


Figure 12: Inaccuracy in traffic forecasts for global rail projects. Data source: Flyvbjerg et al. (2005).

study, and even in possession of it we would be charged with a difficult task if we were to hold the responsible agents accountable for these actions.

3.2.6 Flyvbjerg, Holm and Buhl

Flyvbjerg et al. (2005) did a study of 210 transport infrastructure projects, of which 183 were road and 27 were rail projects. The projects were completed between 1969 and 1998 in 14 different countries, although the vast majority of these are from Europe. Their database has been used in a number of other publications (e.g. Flyvbjerg, Bruzelius, and Rothengatter 2003; Flyvbjerg, Holm, and Buhl 2002, 2006; Flyvbjerg 2007, 2009; Holm 2000). Since these publications all refer to the same dataset, I shall treat them as a single study here. The authors collected data for 31 of these projects themselves, while the remaining 179 are based on data from other studies of forecasting accuracy, including most of the studies mentioned so far in the present chapter.

Flyvbjerg et al. found that both road and rail projects displayed large degrees of inaccuracy, with a trend of large overestimations for rail projects (see Figure 12) and slight underestimations for road projects (see Figure 13). The spread for both types of projects were however very large, with examples of dramatic over- and underestimations for both rail and road projects. Of the 27 rail

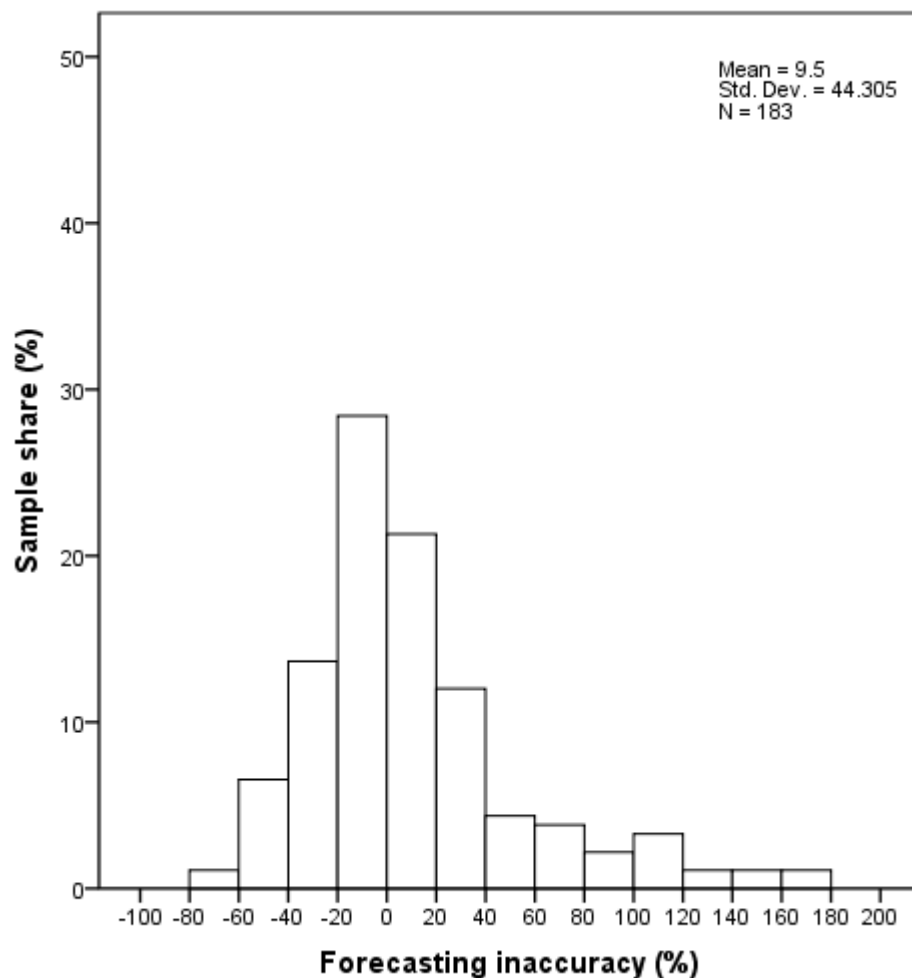


Figure 13: Inaccuracy in traffic forecasts for global road projects. Data source: Flyvbjerg et al. (2005).

projects, two German projects were treated as outliers in the main analysis due to their drastically underestimated ridership forecasts, while the rest of the sample displays a tendency for overestimation. The authors claim that ridership forecasts for rail projects are overestimated by more than 100% on average, which means they must also have excluded the two outliers when forming their conclusions.

The authors also attempt to investigate the dominant causes for the observed inaccuracies in their sample. This is done by asking project managers to state the most likely cause of forecast inaccuracy for projects where the authors collected data themselves, while causes for projects sampled from other studies rely on the statements from the authors of these studies. The main explanations offered for rail projects are trip distribution, deliberately slanted forecasts, and non-specified causes (other). For road projects the main explanations offered are trip generation, land use development, trip distribution and the forecasting model. This has later lead one of the authors (Flyvbjerg 2007) to claim that infrastructure appraisal is governed by a 'survival of the unfittest', since forecasters intentionally understate the costs and overstate the benefits of projects, resulting in the most unfeasible projects being implemented since they appear to provide the best value for money.

The authors claim the study to be the first statistically significant study of traffic forecasts, and have indeed compiled a large pool of road projects to form a distribution of inaccuracies. For this reason it has also been one of the most used references in the field of traffic demand forecasting accuracy, but I should like to add four observations I have made in regard to this study. First, the rail sample is much smaller than the large road sample, and would be difficult to consider statistically significant in any way that differs drastically from earlier studies based on smaller samples. The small sample size is unsurprising given the smaller total population and constricted data access for rail projects, but as many of the claims being made when citing this study are in reference to this particular subset of the sample, such an observation becomes anything but trivial or irrelevant. Second, the vast majority (85%) of the total sample is based on earlier studies, which have then been compiled with data for a smaller set of new projects for which the authors have gathered data. This is in itself not particularly problematic, but it makes it difficult to do a cross-comparison with earlier studies. For example, it is a trivial result that the findings in Flyvbjerg et al. (2005) support those in NAO (1988), Fouracre et al. (1990), and Pickrell (1990), since these studies themselves constitute a large part of the aggregate sample²². The same mention can be made in reference to the stated explanations in the study, since the majority of these are also aggregated from the stated explanations in earlier studies. Third, many of the previous studies included by Flyvbjerg et al. are not able to provide a single cause for forecasting inaccuracy, which makes it difficult to assess the validity of the claim that a “*deliberately slanted forecast*” is the primary explanation of inaccuracy for rail projects (Flyvbjerg, Holm, and Buhl 2005, 139). When multiple explanations are offered, how has the primary explanations for the analysis leading to such a claim been selected? The authors sadly provide little methodological insight to this important question. This is worrisome, since it gives a false sense of statistical support to one type of error, while it remains unclear how these statistics have come about. Finally, the later claims by Flyvbjerg (2007) regarding the survival of the unfittest projects as a result of misrepresented forecasts seem to be largely speculative, as such misrepresentation has not really been thoroughly documented in any of the secondary sources that are referred to. Even in the presence of such documentation the claim would still lack support, as proof of deceptive forecasts for the chosen alternative would not really tell us anything about how valid the forecasts for the dismissed alternatives have been. Such investigations would indeed also be quite speculative, since these dismissed alternatives cannot be compared with any actual outcome for obvious reasons.

As was the case for Pickrell’s study, I do not list these concerns to belittle the efforts of the authors. Even with these four reservations in mind, the study is still among the most important contributions to forecasting inaccuracy in my opinion, as it has provided a necessary focus on a problem that previously received an undeserving lack of attention in planning literature. The study might not provide any particularly convincing evidence of deliberately slanted forecasts being a major explanatory cause, but it does highlight a set of graving concerns about the ability of planners to produce the numbers that best suit their clients. Whether strategic misrepresentation is a major concern in demand forecasts or not, the authors’ recommendations for increased transparency in the preparation of decision support is certainly sound advice, and would likely help to reduce potential misrepresentation as well as other types of uncertainty in the decision-making process.

²² The three studies mentioned here include combined data for 166 road projects and 18 rail projects. However, it is unclear how many of these have been aggregated into the sample of Flyvbjerg et al.

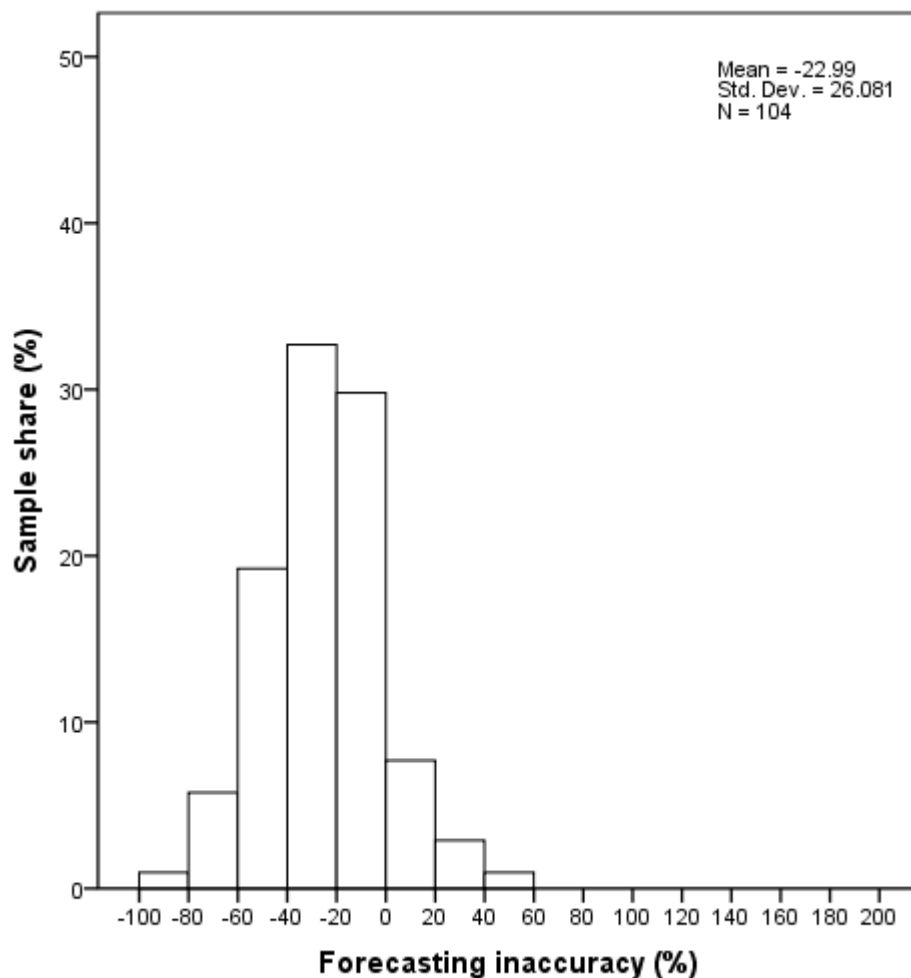


Figure 14: Inaccuracy in traffic forecasts for global toll road projects. Data source: Bain (2009).

3.2.7 Bain

Bain (2009) reports on a set of studies on the accuracy of toll road forecasts he conducted in the period 2002-05 for the credit rating agency Standard and Poor's. The aggregate study includes 104 international toll roads, bridges, and tunnels. Any further specifics on the projects are difficult to come by, as the data set has been completely anonymized. The author claims that this has been done due to the commercial sensitivity of the projects involved. Furthermore, in an e-mail correspondence attempting to shed some light on this, he claims that the supporting data is no longer in his possession and has likely been destroyed, making it practically impossible to obtain further documentation about the sample (Bain 2010). Nonetheless, it remains one of the few comprehensive studies on toll roads, and I have chosen to include it due the implication of the findings.

Although Bain reports inaccuracy in a slightly different format than the other studies in the present chapter, the figures are fairly easy to convert to a comparable format. Doing so reveals a different trend than those observed in e.g. NAO and Flyvbjerg et al., since road traffic now seem to be overestimated rather than underestimated (see Figure 14). Apart from this shift in the observed mean, the distribution of inaccuracy in Bain's study seems to be fairly comparable to the studies on non-toll roads, although with a smaller spread. However, like in earlier studies the spread is still quite

large, with some projects carrying less than 20% of forecasted traffic volumes while others carry more than 150% of forecasted traffic volumes.

In addition to traffic in the opening year, Bain compared forecasts with time series of up to five years after project opening where data allowed for such analysis. These indicate that no improvement in forecasting accuracy appear to be taking place over time, and that ramp up effects have not been able to explain much of the observed forecasting inaccuracy, if any at all. Instead, Bain points to the complexity of tariff systems, overestimation of willingness to pay, and optimistic assumptions regarding economic growth as the main contributors to forecasting inaccuracy. He concludes that optimism is enhanced by the difficulty in assessing consumer behaviour to the direct charging systems of toll roads and that shadow tolls²³ would automatically reduce this potential for error.

The findings are particularly interesting since they display a similar trend as has been observed for rail projects in previous studies, and thus provide solid empirical support for toll road forecasts behaving differently than non-toll road forecasts. This falls well in line with the observations summarized in Flyvbjerg et al. (2005), and indicate that forecasts have a tendency to bias themselves towards figures that would be considered more favourable by its proponents. Bain also points to strategic behaviour as a key source of bias, but provides no evidence for this claim apart from the observation that incentive structures that encourage such behaviour are in place. He leaves it out of his Traffic Risk Index (Bain 2009, 481–482), where he focuses on the many pitfalls involved in forecasting instead, as well as the large span of inaccuracy that covers both grossly underestimated as well as overestimated projects. This spread is typical for all larger studies of demand forecasting accuracy (as shown in Table 1), and even if decision makers were faced with a completely unbiased distribution of inaccuracies, the potential for error would still be huge. Any investor in transport infrastructure should be aware that considerable benefit shortfalls are not uncommon, and that myriads of potential sources of uncertainty exists, of which many could spell financial doom for a project.

3.2.8 Parthasarathi and Levinson

Parthasarathi and Levinson (2010) compared forecast and actual traffic volumes for 108 highway projects in Minnesota that were completed in the period from the 1960s into the 2000s. They seem to have done an extensive collection of data for these projects, allowing them to conduct a linear regression analysis of some of the factors expected to influence forecasting accuracy. Qualitative interviews with a few transport planners were also conducted to shed light on possible causes for inaccuracy. Furthermore, they had access to forecast and count data for individual stretches of the road network, with an average of 25 stretches per project. This made it possible to estimate both average inaccuracies for the projects as well as the span of inaccuracy on individual stretches, as well as categorize projects based on their functional roadway classification.

²³ Shadow tolls are predominantly a UK phenomenon, where the government subsidizes privately-funded roads based on their traffic volume. Bain's logic seems to be that such a system eliminates the direct charge of users, which (*ceteris paribus*) will reduce the complexity of route choice from a user perspective, and thus make it easier to model travel behaviour in the system. I find this a reasonable argument in regards to uncertainty in general, but I am not sure how exactly it would be a cause of bias.

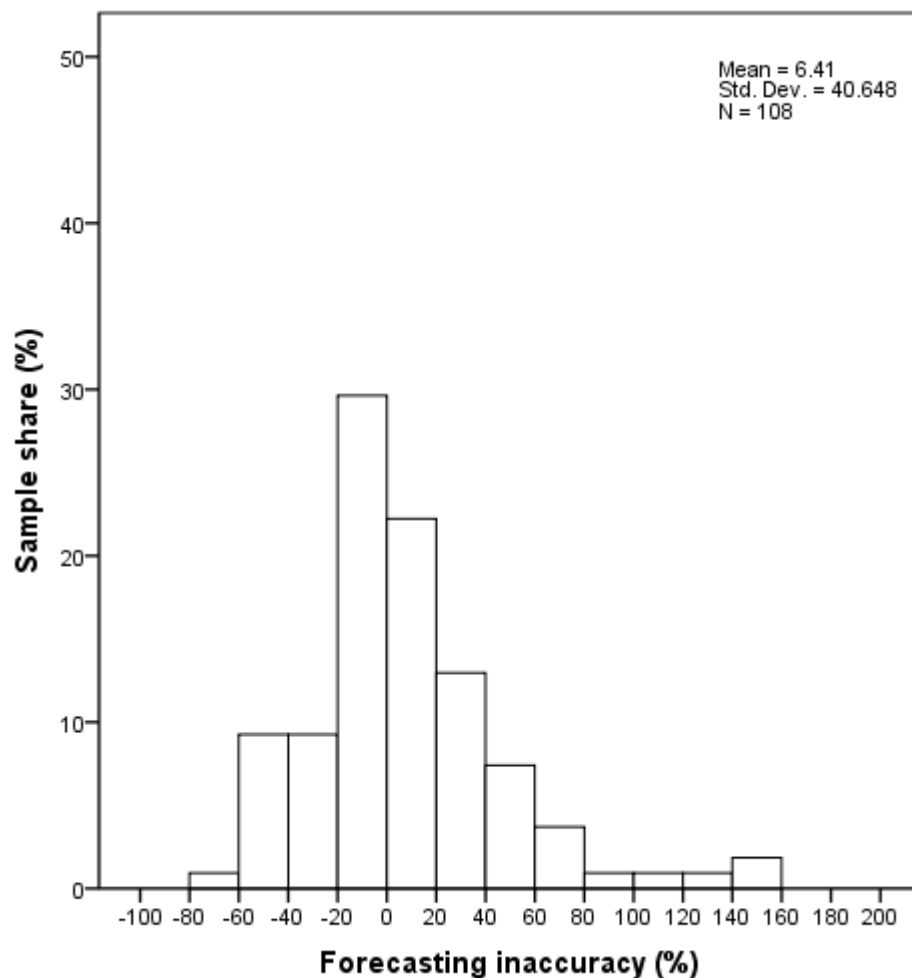


Figure 15: Inaccuracy in traffic forecasts for road projects in Minnesota. Data source: Parthasarathi and Levinson (2010).

The authors provide scatterplots and histograms to give an idea about the distribution of inaccuracy, and report these for a variety of different link types. Across the full span of links there seem to be a slight overestimation of traffic, although the bias is very limited. For critical links however (defined as the links with the most traffic), the trend seems to be a clear underestimation of traffic. Since Mackinder and Evans, NAO, Flyvbjerg et al. and Bain only include primarily links in their analyses it seems fair to assume that the results from these studies should be compared to the highway category in Parthasarathi and Levinson instead of the full range of links. This comparison seem to yield similar results as other non-toll road studies, with a general trend of underestimation for highway links. If inaccuracy for individual projects is measured as an average of link inaccuracy for the project, then we observe a similar trend (see Figure 15).

In the regression analysis it was found that the most influential factors on forecasting inaccuracy were the time horizon of forecasts, roadway classification and the year the forecast had been prepared. The interviews highlighted the inability of models to understand and predict fundamental societal changes as the most important source of inaccuracy. Especially the inclusion of women in the workforce was a common example of unexpected developments that had a major influence on traffic volumes. All interviewees stated that political pressure in Minnesota was probably a less influencing

factor than in other states, and that most criticism from the use of models stems from the inability of politicians to understand the process behind their results.

The findings in the study square well with those of NAO (1988) and Flyvbjerg et al. (2005), and show a slight underestimation of traffic on motorways but also a large spread in the distribution of inaccuracy. However, it also shows that there is a significant overestimation on other link types, which indicate that distributional errors rather than overall traffic volume could be an important source of inaccuracy. This is also reflected in the qualitative data presented by the authors, where most of the respondents voiced concerns over the trip distribution model used in many of the forecasts. Other oft-stated concerns include optimistic assumptions of socio-economic forecasts as well as the models' inability to predict large societal changes. Most respondents agreed that the models themselves were necessary for predictions, but that policy makers often lacked sufficient understanding of the complex processes behind the forecasts and thus ended up applying the results in the wrong manner or to solve a set of problems for which the forecast was not suited to provide decision support. The authors compared some of forecasted socio-economic input variables with the actual outcomes, and found support for the respondents' claim that modellers are often presented with optimistic assumptions about the future growth potential in the region. This is interesting, since this would lead us to expect equally overestimated forecasts for traffic demand, while the converse can be observed in Figure 15. This indicates a model specification error, and lends further credibility to the claim that the models are ill suited to deal with the impact of societal changes on traffic demand. However, without a specific regression analysis that takes these conditions into account, such a claim runs the risk of being based on an ecological fallacy, although in the face of supporting empirical findings this does not seem to be the case here. The authors' claim that these findings call for *"a fundamental rethinking of the meaning, purpose and use of forecasting and modeling methodologies and a move towards adopting a comprehensive view rather than a narrow project related focus"* (Parthasarathi and Levinson 2010, 441) seems to hit at the heart of the problem, since there appear to be too many uncertain assumptions included in the preparation of traffic forecasts to expect any more accuracy than is displayed in the studies presented in Table 1.

3.2.9 Welde and Odeck

Welde and Odeck (2011) did a comparison of 50 Norwegian road projects, of which half were toll roads. This is the only study described in the present chapter that compares forecasting accuracy of toll and non-toll roads explicitly, which allows for direct comparison based on similar data sources. The toll projects were completed in the period 1990-2007, and the project sample thus includes 76% of the Norwegian toll projects implemented during this period. Despite the low amount of projects in this study compared to some of those previously mentioned, the high share of projects still makes the study very representative of the overall Norwegian experience. Where other studies have mentioned a potential bias of data availability, Welde and Odeck found that toll projects actually had quite extensive documentation due to the enhanced data requirements for such projects in parliamentary processes. In contrast, data was difficult to find for non-toll projects, where the sample covers projects in the period 2001-07 for which the necessary data was available.

In line with other studies the authors found that traffic on non-toll roads were underestimated, while traffic on toll roads was slightly overestimated. However, for toll roads the mean is not statistically

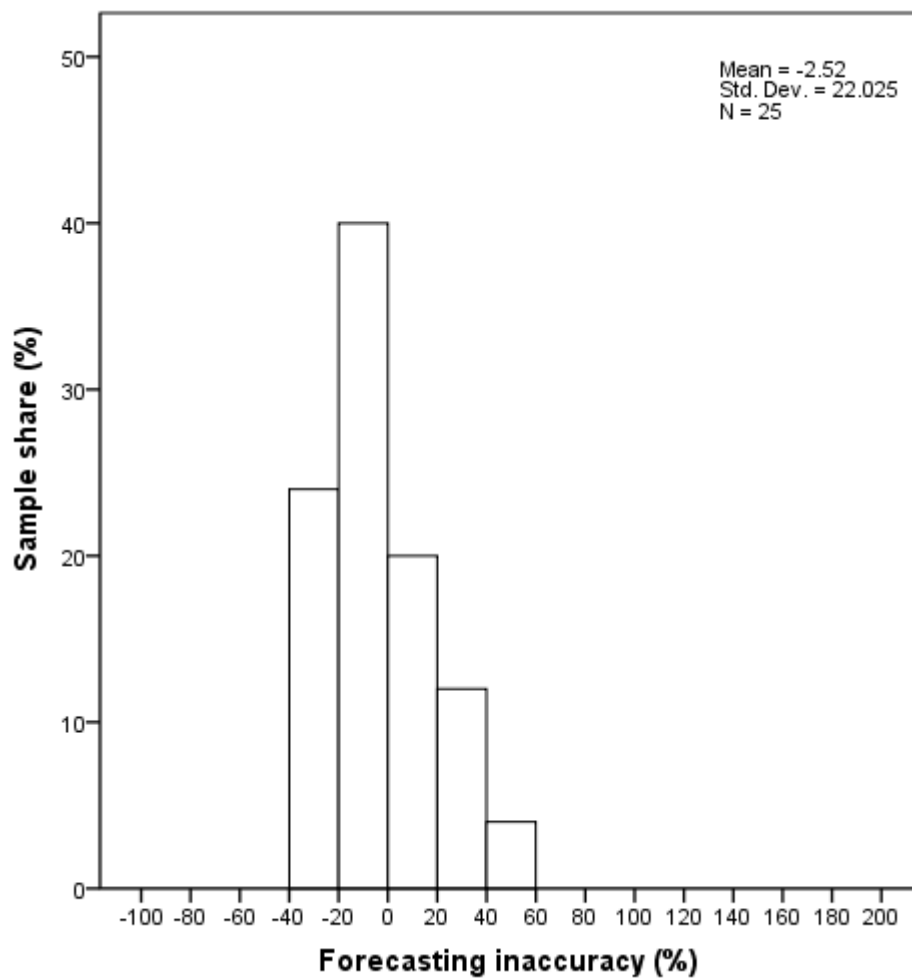


Figure 16: Inaccuracy in traffic forecasts for Norwegian toll projects. Data source: Welde and Odeck (2011).

significantly different from zero, and the observed overestimation could simply be a random error. For non-toll roads the result is certainly statistically significantly different from a zero mean, indicating a general trend of underestimated traffic volumes. The bias is also much larger than the other studies mentioned in the present chapter. However, like all the other studies there is a significant spread of inaccuracy, with many projects experiencing severe under- or overestimation of traffic volumes.

Welde and Odeck also checked for ramp-up effects, and found that for toll-roads the mean inaccuracy changed from negative to positive in the five first years of operation. While not explicitly stated by the authors, we must assume that this result is not statistically significantly different from zero either, but it is still an interesting contrast to Bain, who claims that underperforming projects are unlikely to ever catch up. More extensive research into demand ramp-up for toll projects could probably be a fruitful area of future research. For non-toll roads the underestimation is quite severe. The authors explained this as being a result of the economic growth in Norway during the period in which these projects were completed, and mention induced demand as a possible cause as well.

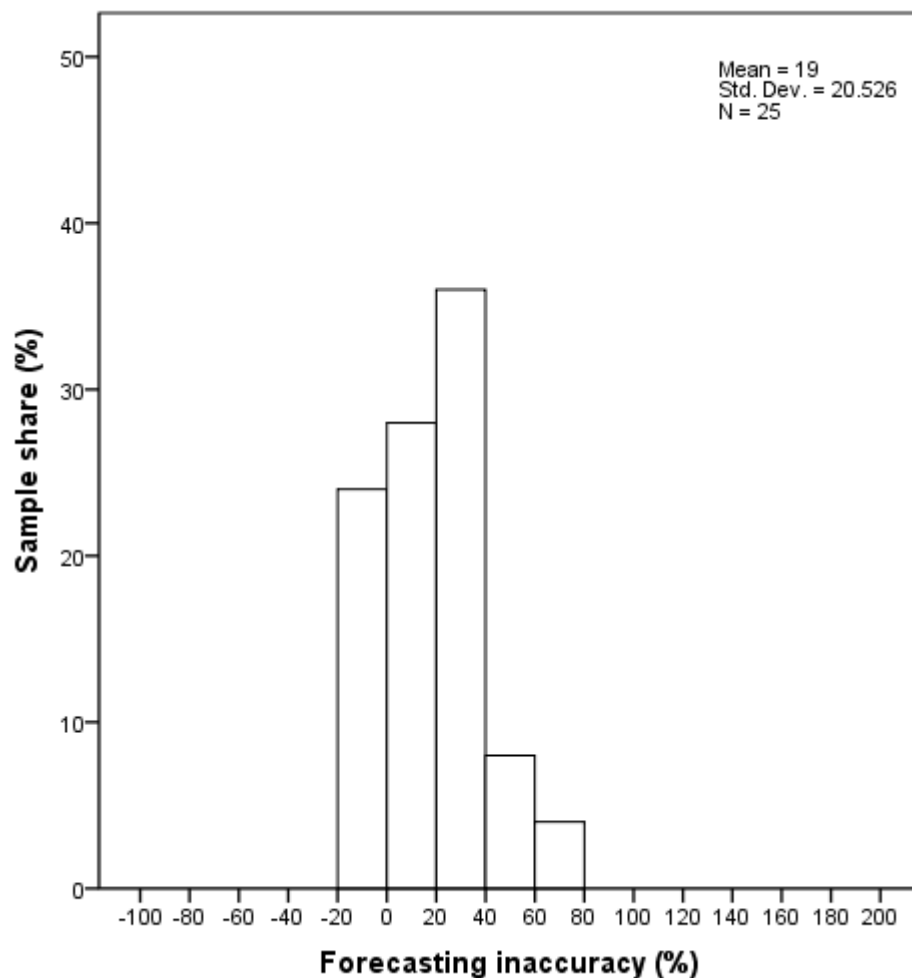


Figure 17: Inaccuracy in traffic forecasts for Norwegian road projects. Data source: Welde and Odeck (2011).

The results for the toll roads indicate similar tendencies as those reported by Bain (2009), although the mean is much less biased this time. However, there is a substantial standard error in Welde and Odeck's dataset, so this might merely be a reflection of the sample size. As has been observed in all other studies presented in this chapter, the spread of inaccuracy is large for both toll and toll-free roads, which in itself is a cause of concern. Moreover, the underestimation of traffic for toll-free roads is very large indeed, and much larger than in any of the other studies I have so far presented. As was the case for the toll sample there is also a large standard error in the non-toll sample, which indicates that the Norwegian projects might not differ that much from the other studies after all, but it is clear that there is a tendency to underestimate traffic for non-toll roads. The authors offer the neglect of induced traffic as an important explanatory cause here, as most Norwegian non-toll projects are evaluated with a traditional four-stage model for transport that uses fixed matrices for trips. In the case of toll roads this does not seem to be the case either, but the effect of tolls is calculated specifically for these projects, and thus account for some of the negatively induced traffic that results from higher travel costs. This could be part of the explanation for why Norwegian toll roads appear less biased than those in Bain's sample, since a good understanding of the elasticity of toll effects have been established in Norway. Indeed, Bain (2009) also mentions that countries in

which toll systems do not have previous local experiences to base these elasticities on are the ones in which planners are most prone to overestimate the traffic demand for such projects.

3.3 Synthesis of findings in prior art

In this section I shall try to synthesize the findings from the studies presented in section 3.2 to form an overview of how best to contribute to the existing body of literature on traffic forecasting inaccuracy. I have tried to split them into general categories that have emerged informally through the process of going through these previous studies, but the categories are mainly to serve as a set of themes that seem to be important in relation to the validity and reliability of traffic forecasts if they are to provide useful decision support to policy makers. The purpose of discussing them here is to serve as a basis for the investigation undertaken in the present study.

3.3.1 Distribution of inaccuracy for different project types

This should be an obvious category to include in any study of forecasting accuracy. As was clearly shown in Table 1, the mean inaccuracy differs greatly between different studies, and particularly between studies of different project types. In addition to this we also observe a large standard deviation for most studies. However, unlike the mean inaccuracy there seem to be no obvious connection between project types and the size of the standard deviation. As mean and spread of normally distributed variables are independent from each other, this is not necessarily particularly odd if we do in fact expect a normal distribution. There is a clear lower bound on inaccuracy due to the way I choose to calculate it, which obviously violates the assumption of a true normal distribution, since inaccuracies can go above 100% (and sometimes do) but never below -100%. As an example of an alternative approach, Salling and Leleur (2006) uses an Erlang distribution for inaccuracies to describe forecasts for construction costs, and since similar bounds exists for inaccuracy measures of cost and traffic this could also be done for traffic forecasts. However, for higher shape values the Erlang distributions converges to a normal distribution, so for all intents and purposes I shall assume the distribution of inaccuracy to be approximately normal, but readers are advised to keep these reservations in mind during the interpretation of results. Alternatively, the inaccuracy could be adjusted through a logarithmic transformation, but like mentioned by Flyvbjerg et al. (2005), this would reduce the intuitive interpretation of the measure, which I have mentioned earlier to be of high priority. The violations of normality described here has little if any impact on the arguments put forward in the present thesis, but could prove important if one attempts to apply the results in quantitative risk assessments.

Under the assumption of normally distributed inaccuracies it is interesting that there are so large differences between samples of the same project type, since we would also expect the spread to be independent of sample size. This is in itself an indication of problems that would arise from attempting to use the results for statistical corrections, since such an approach would involve a considerable amount of uncertainty. If any serious attempts are to be made at this, the available project data would need to be much more detailed than what have been presented in any of the studies listed in section 3.2. As this is not the purpose of the present study, I shall not attempt at correcting for this uncertainty, nor shall I employ any other measures for inaccuracy than I_p . The reason I mention it here is mainly to raise some awareness about the necessary care when employing reference class forecasting or other techniques based on results of ex-post evaluations. Such

adjustments could easily risk making impact appraisals even less transparent without adding much useful information, mainly due to the difficulty in identifying suitable reference classes and obtaining the required detail for data items once these classes have been specified.

3.3.2 Selection criteria for projects to be included

As discussed in further detail in section 3.1, the construction of the dataset can fall victim to quite severe availability bias if sampling is not truly random. Failure to correct for this might result in overly biased samples. Since observed bias has so far been predominantly related to rail projects, I shall briefly discuss this issue only in relation to the studies by Pickrell (1990), Fouracre et al. (1990), and Flyvbjerg et al. (2005), as these are the only studies that include rail projects. Since Pickrell's sample consists of projects decided in a period that saw renewed interest in transit projects in the US, it is possible that the bias is partly due to poor experience among the involved stakeholders with these types of project. Later findings by Button et al. (2010) suggest that the poor forecasting performance reported by Pickrell's original study have actually caused planners to become more aware of the problem, and that forecasts for subsequent rail projects have improved quite significantly in terms of bias (although still with large deviations for individual projects). Similarly, the findings by Fouracre et al. are for metro projects in developing countries, where stakeholders had little or no prior experience with these types of projects. Additionally, it might be difficult to directly compare the institutional settings for infrastructure planning in developing countries with those in Scandinavia for example. Since the findings in Flyvbjerg et al. are partly aggregated through the findings in the two other studies mentioned here, the same concerns naturally also extend to the findings presented in their larger sample. Their study also highlights an inherent problem of aggregating results from other studies of forecasting inaccuracy, as available studies of individual projects might not represent a random sampling of the total population of projects. For example, the inclusion of a source such as Hall (1980) cannot possibly be considered an unbiased sampling approach, since the entire purpose of Hall's work was to investigate notorious planning disasters.

Due to the lack of available data such an approach is understandable and perhaps even necessary for constructing a sizeable dataset. However, such conditions must be kept in mind when inferring conclusions about the representativeness of these studies, since they present quite large violations of the necessary conditions for the inaccuracy measure to be normally distributed. Statistical validity is not only judged by sample size, and it is dangerous to adopt a habit of significance-chasing if this leads to sampling bias. In addition to this, two outliers in the other end of the spectrum were excluded by Flyvbjerg et al. (2005), which enhances the perceived bias. There is nothing inherently wrong in doing so, but certainly these conditions are much greater violations of the assumption of normality than the one discussed in section 3.3.1. In the present study I shall therefore try to include projects based exclusively on collection of primary data sources and omit any inclusion of results from previous studies. The only exception to this principle is sources that can themselves be verified to employ completely random sampling. Furthermore, no projects shall be treated as outliers merely due to the observed inaccuracy of their forecasts. There can of course be good reasons to exclude projects that do not represent the general population well, but surely the accuracy of their forecasts alone cannot be used to determine this, since this is the independent variable we seek to measure. This argument is further enhanced by the fact that a large average deviation in accuracy is the primary motivation for undertaking these types of studies. Statistical outliers are typically caused by

either measurement error, mixing the results of different populations, or a heavy-tailed distribution. Measurement error should be relatively easy to control for in small samples, and exclusion would thus have to be based on special conditions found in the dependent variables instead if we still assume the distribution of inaccuracy to be approximately normal. Since no reason for why this should be the case is presented by Flyvbjerg et al. (2005), I have based the values in Table 1 on the full sample, included the outliers. Note that this does not change the validity of the general arguments put forward by the authors, and it is primarily a methodological concern to be taken into account in the present study. This becomes particularly important in discussing measures to deal with forecasting uncertainty, if it turns out that the outliers are indeed caused by a heavy-tailed distribution.

3.3.3 Structure of transport models

As indicated in several previous studies, the model specification used for individual appraisals could potentially be a large source of uncertainty. The issue of trip distribution and induced demand seem to be the most recurring concerns for road projects, while there is a general concern about the models' ability to deal appropriately with changes in travel behaviour stemming from large societal changes for all project types. However, this limitation is perhaps most crucial for non-routine projects such as larger tunnels, bridges, or rail projects, while many road projects are typically a result of continued trends in private vehicle travel. It is true that effects such as the rapid feminization of the labour market during World War II have had major effects on traffic demand in the US, but for many years transport planning has generally followed a predict-and-provide pattern, where most of the attention for larger transport infrastructure projects has been given to the continued expansion of the motorway network. Apart from major new links between areas that were previously poorly connected, large road projects can safely be assumed to cause a less dramatic impact on travel behaviour than large rail projects. This is hopefully not a controversial statement, as the motivation for large rail projects is often explicitly to change travel behaviour. In the present study I shall try to address some of these modelling issues, but as the previous study clearly show it is difficult to assess these issues in-depth due to lack of available data on model specifications. At best we can hope to get a general understanding of the type of models that have been used in the preparation of traffic forecasts when we are able to define the temporal and spatial conditions for the individual projects.

3.3.4 Explaining inaccuracy and bias

Perhaps the biggest controversy in the existing body of literature on forecasting inaccuracy surrounds the causes that various authors have attributed to the observations. As evident from the findings presented in section 3.2 there are many potential causes of inaccuracy and rarely any conclusive evidence to single out a single cause as a best explanation. Whenever such explanations have been offered it has largely been based on abductive reasoning from the authors themselves. This is problematic for many reasons, of which the most obvious are the risk of confirmation bias and the inability to validate the claims. The exception might be studies with smaller sample sizes (or in the case of Kain (1990), a single project). As these studies typically have much more detailed knowledge of the individual projects, they are also able to provide more convincing arguments for the possible explanations of inaccuracy. However, even when faced with intimate knowledge of the individual cases it is rarely possible to provide more than highly plausible explanations for the observed inaccuracy. Since authors have mainly been looking for confirming evidence rather than

disconfirming evidence, there is still a strong risk of confirmation bias even when detailed accounts are available. Although it was eventually decided to go ahead with the project, Kain's detailed account does of course not leave us with any information of the decision support available at later stages, and thus unable to determine whether this decision was equally ill-informed. Considering that the purpose of the present study is partly to provide empirical evidence for the degree of uncertainty and bias in Scandinavia transport appraisals, it would be naïve to attempt similar detailed accounts for individual projects herein, since the former goal requires a relatively large sample. I shall deal with the issue of explanatory causes in a slightly different manner than any of the previous studies, and a more detailed description of the approach is presented in 4. Of special interest is of course the theory of strategic misrepresentation, due to the moral implications for the planning practice if there is reason to support this as primary cause of inaccuracy, and the difficulties associated with providing empirical evidence for this in the absence of informants by people that are somehow involved in the deceptive actions themselves.

3.3.5 Impact of forecasts on decision-making

One of the things that have struck me the most while going through previous studies of forecasting inaccuracy is the lack of attention to the impact that forecasts have on decision-making processes. Readers are often left with the implicit assumption that policy makers behave in a strictly rational manner, with forecasts of costs and benefits being the sole determinants of which projects get selected for implementation. Obviously this is an overly simplified perception of how decisions actually happen, and there seem to be good reason to investigate this aspect of forecasting inaccuracy more thoroughly. Even if we admit that a perfect correlation between favourable forecasts and project selection is likely not the case, we are still faced with much uncertainty about how inaccuracies in traffic forecasts translate into subsequent appraisal models such as cost benefit analyses and environmental impact assessments. It is not necessarily the case that an unbiased distribution of inaccuracy for traffic forecasts would result in an unbiased distribution of cost benefit ratios, since a forecast that is overestimated by 20% might result in a proportionally larger loss than the benefits resulting from a forecast that is underestimated by 20% (or vice versa). In addition to this the impact of forecasting inaccuracy for different project types might also differ. For example, some of the projects in Fouracre et al. (1990) are considered quite successful in spite of gravely overestimated ridership figures, while NAO (1988) point to large inefficiencies in the allocation of funds as a result of higher traffic volumes than expected. I shall try to address both of these limitations in the present study. An important issue that seems to have been overlooked in all previous studies of forecasting inaccuracy is the comparison with the zero-alternative, and the accuracy of forecasts for this scenario. This is perhaps understandable, given that focus has been on establishing samples of completed projects, but nonetheless an important shortcoming that it would be fruitful to investigate in the present study. Næss (2011) argues that this is an important source of bias in appraisal for road projects in particular.

3.4 Summary

The purpose of this chapter has been to provide an overview of some of the most important previous work that has been done on the accuracy of forecasting demand for transport infrastructure projects. Previous findings in relation to the accuracy of cost estimates was also addressed, but with references to recent reviews conducted by other authors rather than a comprehensive review as for

studies of demand forecasts. The review displayed a noticeably degree of forecasting bias in many studies, especially for studies focused on rail projects. Furthermore, the standard deviation differed quite considerably between studies, but was fairly high in all of them. Most importantly, the review highlighted several theoretical and practical problems in the interpretation and comparison of findings from different studies, which are also highly relevant for the analysis in the present study.

4 Methodology

Traditional scientific method has always been at the very best, 20 - 20 hindsight. It's good for seeing where you've been. It's good for testing the truth of what you think you know, but it can't tell you where you ought to go.

- Robert Maynard Pirsig

In this chapter I will elaborate on the core problems addressed in the present study and explain my choice of methods to examine them. This includes a discussion of suitable units of observation as well as the data sources necessary to obtain these units. I shall also briefly cover the rationale behind the methods employed in the present study as well as the opportunities and limitations that are embedded in this selection. The discussion presented here is to serve as guidelines for the empirical analysis, but in practice it might of course not be possible to follow these to the letter at all times. Where compromises have had to be made, this will be explicitly stated in the presentation of data in the next chapter.

4.1 Addressing the problem statement

Let me start this chapter by recalling the problem statement of the present study as it was described at the end of chapter 1:

How accurate are traffic forecasts used for ex-ante appraisals of transport infrastructure projects, and how does this accuracy affect the decision-making process?

Clearly the above statement and associated research questions are still quite broad, and could themselves be branched into entire studies. Further demarcations are thus necessary to allow a fruitful inquiry of value for planning theory and practice. I shall attempt to make such demarcations explicit throughout the present chapter, and for the sake of simplicity I shall attempt to discuss the methodological considerations for each research question separately. There are undoubtedly many overlapping features between these questions and the empirical investigations available to us that speak against separating these in practice, and so it is mainly a choice I have made in regard to the pedagogic effect of such a structure as it will be presented here.

4.2 Degrees of inaccuracy in forecasting

Let us recall the first research question of the present study:

1. **How well do forecasts presented in decision support for transport infrastructure projects match the actual development?**
 - a. What is the extent of forecasting inaccuracy in transport project evaluation?
 - b. Is it possible to identify characteristics of projects that are prone to forecasting inaccuracy?
 - c. What is the influence of forecasting inaccuracy on subsequent impact assessments?

Observed inaccuracies in previous studies, of which some have been presented in chapter 3, remain the underlying motivation for the present study, so mapping out uncertainties for the relevant case areas must be a necessary point of departure for studying the effect on decision-making processes. After all, studies based on larger samples are rare in this field, and none of them capture the case areas of the present study. Some are too limited in capture (NAO 1988; Welde and Odeck 2011), others are too broad in capture (Flyvbjerg, Holm, and Buhl 2005), or have aged too much to be considered representative of state-of-the-art empirical findings (NAO 1988). There are thus good reasons to collect new data on the degree of inaccuracy of traffic forecasts in Scandinavia and the UK for the present study. The extent of the inaccuracy (question 1a) refers to the statistical distribution of observed inaccuracy, while identification of projects prone to inaccuracy (question 1b) refers to characteristics that might be common among projects that display a high observed inaccuracy. As we have already seen in chapter 3, overall project type (road/rail) might be a good indicator for the degree of bias, and it would be desirable to check for a range of other variables such as the period in which the project opened, the time between the prepared forecast and the opening of the project, possible cost overruns or delays, etc.

How then, having settled that the information is necessary for the study, should we go about measuring the inaccuracy of traffic forecasts? The most straightforward way of doing this would be to use the formal definition of inaccuracy from section 2.4 that was used for most of the results presented in chapter 3²⁴:

$$I_p = \frac{O - P}{P} \times 100$$

While this measure might seem relatively simple, there are many theoretical and practical obstacles in employing it. For example, what constitutes a project for which we wish to measure inaccuracy? A piece of new highway might be planned as multiple sections, which are supposed to be constructed sequentially over a period of 10 years. However, prior to deciding to build the highway a complete appraisal of the full effects of all combined stretches is usually made. Along the way there might be new assessments of subsequent phases as prior phases near completion, in order to update the available knowledge in light of new discoveries or changes in development since the original decision support was prepared. Is the project we wish to assess then the complete highway, or is each individual stretch also a project? In case of the former, how do we deal with potentially major design changes later in the process, such as individual sections being scrapped or a bridge section being expanded from two to three lanes? In case of the latter, do we use the initial forecast for the entire highway, or a subsequent forecast tailored for the specific section? Surely a forecast cannot be expected to be accurate for a different network design than what was used in the original assessment, especially not if some of the planned sections are never built. If we expect an AADT of 40,000 on a link that is later scrapped, is the inaccuracy then -100%?

In addition to defining the project we also need to decide how much of the network we include in the appraisal. Is it only traffic on the main link that we are interested in, or do we also include traffic on the existing network? In case of the former, how do we know whether large inaccuracies are not just

²⁴ P: predicted value, O: observed value

caused by a fairly accurate forecast for overall demand, but with a wrong expectation for how it would distribute in the network? In case of the latter, how do we decide which parts of the network that is relevant to our appraisal? Furthermore, if we stick with just the new link, do we chose a single point of measurement or do we aggregate an average uncertainty based on multiple point estimates?

Given that we find a suitable solution to the above problems, how do we decide which forecast to use if multiple assessments have been carried out in the decision-making process? Sticking with the original forecast might cause the data to be horribly outdated for the last sections if construction takes place over a long period of time. Using a later forecast involves path dependencies of prior commitment and sunk costs, which might result in more accurate forecasts, but perhaps some that had relatively minor impact on the political decision on whether to implement the project. It would be irrational to decide against the final section of a highway that were to link regions that were previously poorly connected no matter what the forecasts might say, if this section was only a minor part of a larger project in which all other sections had already been completed. This issue is certainly not negligible. For example, Brinkman (2003) notes that the Tyneside Metro has been evaluated in four different studies, of which two concluded that forecasts were fairly accurate (P. Hall and Hass-Klau 1985; Mackett and Edwards 1998), while the other two concluded that forecasts were overestimated by roughly 100% (Fullerton and Openshaw 1985; Walmsley and Pickett 1992)²⁵. Brinkman criticises the former two studies for lack of proper sourcing, and claims that since it is impossible to evaluate the validity of them we should trust the studies that conclude forecasts to be highly overestimated. It is true that sourcing is a problem for Hall and Klau²⁶, but Fullerton and Openshaw provide sources that allow us to conclude that their assessment was based on 1982 data for a project that was only partially opened at the time. Data from Walmsley and Pickett indicate that ridership figures roughly doubled from 1982 to 1985, and so it is not unreasonable to expect that all four studies of the project are in fact correct in their assessment of forecasting inaccuracy. Ridership was well below expectations for the partially completed metro, then rose to match expected figures upon completion, and later declined due to increased fares as a result of deregulation. However, while all studies of the accuracy might be correct, it highlights the difficulty of assessing inaccuracy via point estimates.

Turning to the actual count data to represent traffic levels after project completion, we are faced with problems of finding a suitable reference year. Typically this is done by selecting the opening

²⁵ In practice there are only two substantially different studies, since data in Mackett and Edwards is based directly on Hall and Klau, while data in Walmsley and Pickett is based directly on Fullerton and Openshaw.

²⁶ The case is much more complex than I have time to discuss here, but I should like to add a few clarifying notes. The problem of sourcing is not that sources are not specified in Hall and Hass-Klaus, as Brinkman claims. Rather, the problem is that the ridership forecasts they use are based on a report published after the project was approved. This report might itself refer to the original estimates, but this has not been possible for me to verify. If this is not the case, the comparison is based on a different project stage, and the results are thus not in conflict with Fullerton and Openshaw. Conversely, Fullerton and Openshaw refers to an unpublished report as a source of observed ridership, which makes it difficult to verify the figures. To add further confusion, Hall and Hass-Klaus compare annual ridership figures, while Fullerton and Openshaw compare peak hour loads. Both studies are examples of how difficult it is to obtain the required data for these types of comparisons, and how meticulously sources need to be reported if there is to be no confusion about the results.

year, as this is usually also the target of the forecasts presented in decision support documents. For example, Flyvbjerg (2005) has argued heavily in favour of using the opening year (or first year of operations) as a reference point, while acknowledging several points of critique from planners against this methodology. Potential ramp up effects is cited by Flyvbjerg as the most typical objection, which he seemingly refutes by citing a number of cases where ramp up failed to materialize. However, such anecdotal evidence must be used carefully, as there could be many reasons why ramp up does not seem to materialize. As a possible, but equally anecdotal, counter example, we can look to the New Little Belt Bridge in Denmark, which was completed in 1970. At that time it carried 20% less traffic than expected, which worsened to 50% below forecasts 10 years past opening. While traffic in the opening year was low in this case, it would be erroneous to conclude a lack of ramp up based on the traffic level 10 years past opening, as the forecast for this year assumed the Great Belt Bridge to be completed at this time. However, this project was not completed until 1998, and in 2000 (30 years after opening) traffic on the New Little Belt Bridge exceeded forecasts by some 30%. Today the traffic levels are beyond the expected peak demand in the lifespan of the bridge. Had the Great Belt Bridge been completed in 1980 as assumed in the forecasts, traffic on the New Little Belt Bridge would most likely not have trailed behind forecasts by such drastic figures, and increasing traffic levels would then indicate that ramp up could very well have taken place in the years after project completion.

Flyvbjerg thus raises an important issue when he calls for ramp up effects to be empirically documented in decision support, but it might not be possible to quantify the uncertainty associated with this. Given that he justifies exclusion of ramp up effects in ex-post appraisals due to data availability, the same argument must hold for planners that prepare forecasts for decision-making. If it is not possible to study the ramp up effects in isolation based on the available data, how are planners then supposed to provide accurate assessments of such effects? If we allow researchers to demarcate themselves from these issues, I see no reason not to grant the same freedom to planners. This does of course not remove the uncertainty related to such demarcation, and it would appear that more systematic data collection for ex-post evaluation needs to be institutionalized if we are to make better assessments of ramp up effects. That being said, I agree wholeheartedly with Flyvbjerg's argument for simplifying appraisal methods as much as possible, to allow for methodological transparency as well as an adequate sample size, since data availability appear to be the most limiting factor for appraisal of forecasting inaccuracy. It is also important that planners then highlight these uncertainties in the decision support prepared for policy makers, if they are indeed aware of ramp up effects but unable to quantify them in a satisfying way. In cases where planners present their forecasts with pinpoint accuracy, Flyvbjerg is correct in his criticism of their failure to account for ramp up effects.

The issues I have raised so far are by no means trivial, and neither do they constitute an exhaustive list of considerations to be made for assessing the inaccuracy of traffic forecasts. However, I feel that a sufficient amount of issues have been raised to justify a choice of demarcation in regard to the first research question, for the simple reason that many of these are bounded by access to data. For example, a solution to the choice of which parts of the network to include could be to use several screenlines across the main transport corridor for the project we are interested in analysing. This allows us to capture potential inaccuracy arising from network distribution in a simplified manner. An

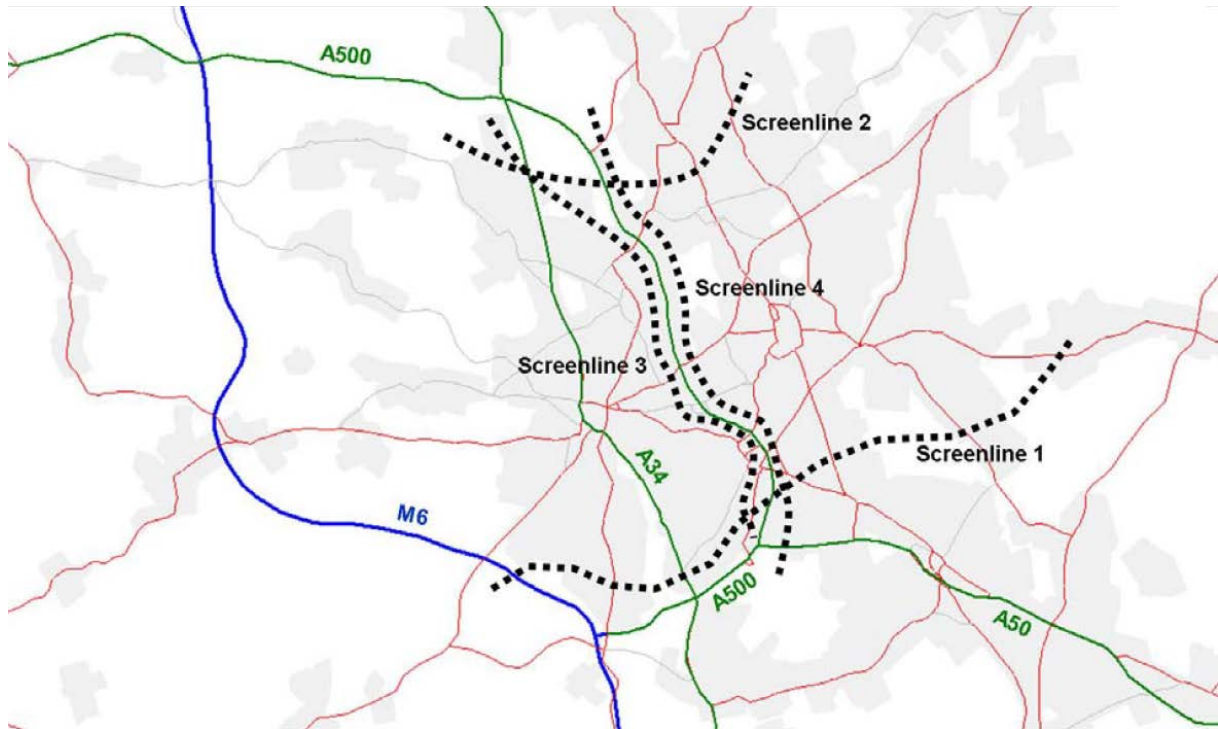


Figure 18: An example of multiple screenlines to account for distributional effects in a larger transport network. The example here is from a junction improvement scheme around Stoke (UK). Reprint from ATKINS (2009).

example of this is presented in Figure 18, where four screenlines are used to cover the main transport corridors affected by improvements to the A500 motorway near Stoke in the United Kingdom. However, this requires that traffic levels for all of the affected roads are available consistently in both decision support and ex-post count data, which is rarely the case. It might be possible to uncover such data if one was determined enough, but such a process could be immensely resource demanding if the data is not readily available. For the present study I doubt that the exponential increase in required resources to uncover the data would be justified by the increase in analytic validity. It would seem that the authors of the related works presented in chapter 3 have arrived at a similar conclusion, since comprehensive screenline analyses are not present in any of them.

With the above considerations in mind, I specify the following overall guidelines for how to collect the necessary data to answer the first research question:

- **Project specification:** Individual projects are considered to be those that are specified in the decision support at the time of political decision to implement it. If the decision support specifies multiple projects or several stages of a project to be completed individually, these are also considered individual projects. If some of these projects are shelved for an extended period of time, e.g. due to lack of public finances, and the forecast was based on completion of the full project package, none of the related projects will be included in the analysis. In such cases it would be difficult to argue that the comparison is a valid appraisal of the expected traffic volumes. It is of course largely a subjective effort to discern whether individual projects will have a substantial effect on each other in a project package, and such

cases have therefore been discussed both within the UNITE team as well as with external stakeholders associated with the affected projects.

- **Reference point for decisions:** Forecasts are selected from the supporting documentation that was available when a political decision to implement the project was taken. Typically this will be in the form of a CBA/EIA or similar types of documentation for individual projects. Figure 19 displays the typical phases of an infrastructure project, in which we are interested in the documentation that is prepared for phase 2. However, this is not always a trivial task, since there is usually a considerable lag between the preparation of decision documents and the opening year of projects, both due to long construction times of larger projects as well as delays of various kinds. The decision support for projects where both forecasts and counts are available will thus often be rather old, and it might be problematic to gain access to the original documents.
- **Reference point for counts:** Estimated traffic figures are based on counts from the databases of the respective administrative body for each project, and based on the first full year in which the project has been in use (i.e. projects opened in 1994 will be based on count data from 1995). Using data from the opening year would require a breakdown of traffic counts before and after project opening, which is not always available. Using the first year after opening as a reference point should do away with most uncertainties of this kind, as well as potential sightseeing traffic immediately after opening of larger projects. In the case of a larger project being split into several stages, counts will typically be based on the first full year of operation after completion of the final stage. Forecasts for such projects are rarely detailed enough to report expected effects of individual stages, so even when operation has begun on the first stages and counts are available, it would provide very biased results if these were compared with forecasts for the full project with all stages completed. Uncompleted stages could act as a bottleneck for traffic to and from completed stages, and in such cases traffic would likely fall well below forecasted values on completed stages.

The reference points and project demarcations listed above are of course ideal types, but in practice it is rarely possible to create a data set for each project that fulfils all of these. A particularly prominent problem is that of mismatching reference points for forecasts and counts. For example, a forecast might be prepared in 1980 that uses 1985 as a reference year for expected traffic volumes after project completion. The project might gain political acceptance in 1982, start construction in 1986, and stand completed in 1993. It would thus be misleading to compare the 1985 forecast with the 1994 count data (one year after opening), as a general growth trend is typically assumed in such forecasts. Failure to account for this would likely make the comparison appear as if the forecast had severely underestimated traffic, if we assume that the prediction of other development trends was correct. Had the forecast been prepared with 1994 rather than 1985 as a target year, it would have taken into account the assumed growth trend in the 9 year gap. Certain authors will argue that such bias should be explained with a control variable for discrepancy between expected and actual opening years, since this is a case of a delayed project. However, forecasts are often prepared for a standard reference year that need not correspond to an expected opening year, and especially so for

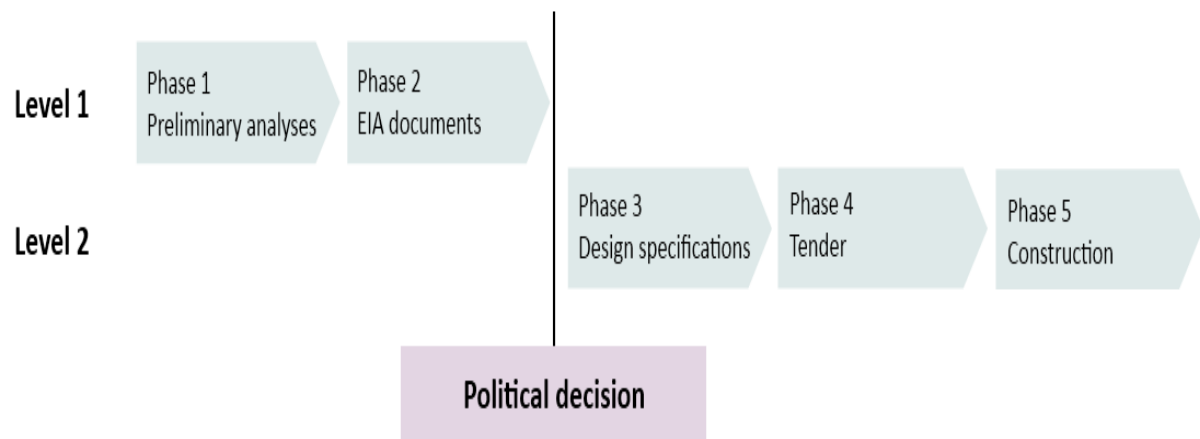


Figure 19: Decision process for infrastructure projects. Reprint from IK (2008), translated by the author.

projects that have yet to reach political consensus. Forecasts for some projects have been prepared with the target year set to the same year that the forecast was made, simply to illustrate the resulting effects in a contemporary situation that decision makers can easily relate to. In such cases the target year for the forecast is clearly not intended to indicate an expected opening year for the project. In other cases a series of forecasts for different projects with the same target year might be prepared simply because it saves a considerable amount of resources. Target years are just a useful reference point in these cases, and expected opening years might therefore differ.

Due to the abovementioned reasons I have therefore chosen to adjust forecasts for individual projects based on the expected growth trend presented in the decision support. I have decided against using observed growth trends in the period for which to adjust, since the goal of the adjustment is to represent the available knowledge ex-ante rather than ex-post. If no growth trend is listed in the decision support documents, I have used 1.5% as a default annual growth trend, which seems to fall well in line with the figures from projects that explicitly list the assumed growth factor. Unless data availability is extremely problematic, many projects will require only minor adjustments or none at all, since data fit the desirable reference years in a satisfactory way. For those that do not, it will not be of great importance to the comparison whether the growth rate is set to e.g. 1.5% or 1.8%. The important thing is that an exponential growth is taken into account for projects where the mismatch of reference points is considerable, in order to do away with most of the bias that would otherwise be present in the results. The adjustments have been carried out by use of the following formula²⁷:

$$A = O(1 + r)^{Y-T+1}$$

As should be obvious from this formula, observed values are adjusted to the first year of operations in case of mismatching reference points. It also allows for adjustment of count data in cases where such data is not available in the desired reference year, but where data might be available for later years (by swapping the forecast target year with the counting year). In cases where the reference points for forecasts and counts match each other but not the opening year (e.g. forecast target and

²⁷ A= adjusted value, O=observed value, r = growth trend, Y = opening year, T = forecast target year

count year is 2010 but opening year is 2007) the adjustment will have no influence on the measured inaccuracy, as both figures are adjusted with the same factor, and we are mainly interested in the ratio between them.

All of what has so far been described in the present chapter has dealt with completed projects and their effects on travel behaviour, and how good planners have been at anticipating these effects. However, in order to arrive at useful results for EIAs and CBAs, the chosen alternatives must be compared with a zero-alternative, in which the expected consequences of not building the projects are assessed. The zero-alternative forms the basis of comparison for which the benefits of building the project is calculated, including travel time savings, noise levels, accidents emissions, etc. In chapter 2 I discussed events as a result of mechanisms being triggered, and like any forecast the zero-alternative is an assessment of likely events if certain mechanisms are triggered. However, most research so far has been focused on the validity of forecasts for completed projects, and assessments of the validity of forecasts for the zero-alternative is thus largely limited to counterfactual arguments, since we have not observed the corresponding event of mechanisms that didn't trigger (i.e. how traffic would have developed if the project had not been built). It would thus also be desirable to investigate the possible inaccuracy of projects that were not completed in cases where the necessary data is available. Sadly, this is not always the case, since forecasts for the zero-alternative are not always explicitly expressed in decision support documents. This seems to have improved in more recent project appraisals, but since we are forced to base our comparisons on older forecast, this is a problem for the majority of the projects available to the present study. For projects that have been completed, it will sometimes be possible to use the above adjustment formula for mismatching reference years to interpolate the forecast for the zero-alternative to a reference point prior to construction start, and use this as a suitable proxy for the validity of the zero alternative.

In relation to question 1c it would probably be most prudent to investigate the quantified outcomes of subsequent impact assessments when adjusting the input variables to reflect observed imprecision or bias in transport forecasting. However, it is highly impractical to have to reconstruct the impact assessments for a large sample of completed projects in light of new knowledge, and in most cases it would be downright impossible to acquire the necessary information to allow this. Many of the implications also seem fairly straightforward, such as ridership shortfalls resulting in a loss of expected revenues for a transit service provider, and while it would be of interest to quantify the effects of this by creating a large sample of reassessed projects it would be undesirably resource demanding to do so, when much of this can be accomplished with logic reasoning.

Given this resource restriction, it seems more fruitful to use a case-study approach and investigate the results of critical cases. At first glance the findings in the related works presented in chapter 3 indicate that forecasting problems are more severe for rail projects than for road projects²⁸. As critical cases to challenge these perceptions it could thus be of interest to investigate whether severe ridership forecasts always ought to be considered problematic bias for rail projects, and whether slight underestimations of traffic volumes always ought to be considered a negligible bias for road projects. However, in the present study we shall mainly address the latter of these situations since the available data from related works presented in chapter 3 clearly indicate that road projects seem

²⁸ With the possible exception of toll roads (Bain 2009; Welde and Odeck 2011).

to provide a more representative set of results for the accuracy of traffic forecasts than do rail projects. Apart from projects completed in a period of stagflation, all studies of forecasting accuracy for roads converge towards a slight underestimation of traffic levels. The data for road projects thus seem to be more conclusive than that of rail projects. While the bias for rail projects is relatively large compared to road projects, it is difficult to gauge whether this is merely due to small sample size for this type of projects. The present study is partly aiming to address this lack of knowledge, but for the subsequent impact assessments we shall make do with logical inferences rather than quantitative modelling.

It might seem counterintuitive to use representativeness of previous results as a selection criterion for a critical case study, as the entire purpose of a critical case is to highlight issues for which we do not need a large sample. However, the critical part in this context is related to the implications for the subsequent impact assessments, and not to the generalization of bias in traffic forecasting. We are thus interested in investigating whether slight underestimations of traffic for road projects can exercise more severe influence on the validity of decision support beyond the seemingly negligible bias observed in previous studies, if we assume that slight underestimations of traffic appears to be a general trend in forecasts of road traffic. Several of the studies presented in chapter 3 attribute underestimated road traffic to the effect of induced travel (Fouracre, Allport, and Thomson 1990; NAO 1988; Welde and Odeck 2011), and if the increase in traffic volumes results in congestion it can cause benefit shortfalls and increased costs (Department for Transport 2006; Marte 2003; Nielsen and Fosgerau 2005). A study of the effects of induced traffic on the impact assessment for a road project in a congested network could therefore act as a critical case. If slight inaccuracies of traffic forecasting do not have significant effect on subsequent impact assessments here, it is unlikely to be a problem for project appraisals in general. This mainly alludes to observed bias in traffic forecasting. Furthermore, if slight inaccuracies of traffic forecasting indeed have a significant effect on subsequent impact assessments, the problem must certainly be more severe for projects with large inaccuracies²⁹. This mainly alludes to observed imprecision in traffic forecasting.

4.3 Planning in the face of uncertainty

Let us recall the second research question of the present study:

2. How is uncertainty managed in the decision-making process?

- a. How are uncertainties communicated between stakeholders?
- b. What are the typical explanations offered for inaccurate traffic forecasts?
- c. Which obstacles do stakeholders identify in the use of traffic forecasts?

As I discussed in chapter 2, there is a variety of uncertainties that pertain to model-based decision support such as those used in demand forecasting. The target of the analysis in the first set of research questions can be considered the aggregate impact that these uncertainties have on the models' outcome, i.e. the accuracy of forecasts and impact on subsequent assessments. However, this does not in itself provide much insight on the influence that these uncertainties exercise on the decision-making process. The focus of the second set of research question is to address this issue by

²⁹ Even if slight inaccuracies turn out to have negligible effects, large inaccuracies might of course still be a problem.

investigating how uncertainties are dealt with by decision makers. After all, if stakeholders have full awareness of the uncertain nature of model-based decision support, the observed inaccuracies between forecast and actual traffic volumes might not pose a great problem for the overall validity and reliability of decision support in transport infrastructure planning. Such awareness (or lack thereof) is crucial to informed decision-making, as the technical analyses might otherwise exercise a large influence on the scope of initiatives that are perceived as suitable to address a given problem. Proper knowledge of uncertainty does of course not exclude the option for policy makers to ignore or overrule model-based decision support based on other considerations, and so we cannot simply use the comparison of forecasts and observed traffic to investigate this issue.

Communication of uncertainties can be done in many ways, and I shall obviously need to limit my scope of inquiry to specific forms of communication for the present study. The goal is not to provide an exhaustive study of possible channels of information flow in policy settings, but rather to investigate whether uncertainties and inaccuracies are dealt with systematically through the decision-making process. An obvious source of information on structured communication presents itself in the form of the decision support that is prepared for policy makers. Indeed, the entire purpose of EIA and CBA guidelines are to ensure a structured assessment of the impacts of a proposed project scheme, which serves both as decision support for policy makers and as justification of project approval or rebuttal towards the general public. It would thus be the most logical forum in which to present uncertainties related to the assessments in a structured manner as well. The official EIA review guidance for projects in the EU explicitly states that uncertainties relating to data collection, survey methods, impact assessment methods, mitigation measures, and a non-technical summary should be made explicit and discussed in the decision support prepared for policy makers (EC 2001). It would thus be logical to assume that policy makers expect the most important uncertainties to be addressed in the official decision support documentation.

Using the EIA documents as a reference point we are thus able to form a proxy of the communication of uncertainties between stakeholders; in this case between planners and policy makers. The decision support prepared for policy makers is of course also dependent on a range of studies included in sub-reports, each of which entailing its own specific set of uncertainties. It would be fruitful to track the communication of uncertainties in these prior steps as well, but I have decided against this mainly due to the following four reasons. First, it would be exceptionally time consuming to do so, since there is often a broad range of analyses involved. Second, the official decision support documentation is already problematic to uncover for older projects, and naturally the availability of more technical sub-reports is even more problematic. Third, many of the analyses might never end up in a written account, but instead take place as an informal trial-and-error approach with consultation directly between stakeholders. Fourth, even if these analyses were readily available it is highly unlikely that policy makers have the time or inclination to go through all the sub-reports prepared for a project. It thus seems reasonable to expect that the uncertainties deemed relevant to policy makers are included in the main reports. Given that the decision support documents prepared for policy makers is already required for the analysis related to the first research question covered in section 4.2, it seems both a reasonable and effective approach to base part of the analysis related to the second research question on these sources as well.

In relation to the influence of uncertainties in traffic forecasting we are of course mainly interested in the communication of uncertainties that relate to the volume of expected traffic levels, as well as the effects that these uncertainties have on the assessment of economic, environmental, and social impacts of transport infrastructure projects. The units of observation can then be chosen as the range of possible outcomes that are provided for the presented traffic volumes, as well as the conditions specified for the forecast to be representative of future development. If a forecast is presented as a point estimate, there is only limited communication of uncertainty. If a forecast is presented as a span of values that future traffic volumes are expected to fall within, this would be an acknowledgement of greater uncertainty in the results. Likewise, if no particular specification is provided of necessary circumstances under which this traffic would materialize there is only limited communication of uncertainty. In a worst case scenario of limited communication of uncertainty, policy makers might consider the expected development of traffic levels as determined by factors that lie out of their sphere of influence.

However, even in absence of an uncertainty span there might be more qualitative descriptions of uncertainty, and I have therefore tried to form a practical measure for uncertainty communication based on the framework described in section 2.3. As this framework is intended as a conceptual rather than practical frame of reference, the measure I employ is necessarily a rather crude simplification. The communication of uncertainties will initially be classified according to their location and then ranked according to their level in accordance with the categories specified in section 2.3. The nature of uncertainty will not be included in this measure but instead assessed qualitatively. There are two main reasons for this. First, it is impractical for the quantitative analysis to include too many dimensions, since this would require a great deal of variables. Second, the framework described in section 2.3 is meant as comprehensive classification. It is unlikely that all combinations of categories are equally relevant, and even more unlikely that they are reported explicitly in decision support documents. In this regard, the nature dimension is easily the least relevant for the present analysis and thus the most suitable for exclusion. In the extent that it is deemed relevant, I shall address this uncertainty dimension when summarizing the findings in chapter 8.

Although there are now formal requirements to report uncertainties in decision support documents such as EIA, it would be naïve to think that stakeholders have previously been oblivious to uncertainties in model-based decision support if they were not made explicitly aware of these in formal documentation. It would thus be interesting to supplement the findings from the decision support documents with a more in-depth inquiry into how uncertainties enter the more informal communication that takes place between stakeholders. In this regard it would be desirable to observe these exchanges at first hand by means of direct observation, but there are obvious limitations to such an approach, not least of which is gaining access to negotiations between stakeholders. Even if this can be achieved there is no guarantee that a neutral observer will not influence the behaviour of observed parties, and direct observations might not reflect a representative communication of uncertainties. Direct participant observation typically requires the observer to become a natural part of the environment in which to conduct such observations, and properly doing so would likely necessitate that this was the sole purpose of the present study due to the time required to build a relationship of trust with the observed parties. I was fortunate enough to

be invited as a participant in several meetings between Odense municipality and COWI as they were scoping the concepts for an integrated model to cover both transport and environmental effects of new policy initiatives. While they could possibly be considered a form of direct participant observation, I have not included any experiences from them directly in the analysis. Instead, they have helped shape the analysis of other sources through the tacit knowledge gained from these events.

Alternative approaches to studying the informal communication of uncertainties are quantitative approaches in the form of questionnaires and qualitative approaches in the form of interviews, or quasi-qualitative/quasi-quantitative combinations hereof. Closed-end questionnaires could provide quantitative measures for the communication of uncertainty, while semi-structured interviews could provide qualitative measures of it. The interviews are perhaps the most suitable in this regard, in order to allow for an exploratory approach to the formation of key concepts in the communication of uncertainty. The uncertainty framework outlined in section 2.3 could then act as a structural categorization scheme for the emerging concepts in the analysis of the interviews.

In relation to identification of explanations offered by stakeholders it seems that questionnaires and interviews with stakeholders would also be the most suitable approach. A point of departure can be taken with inspiration from the related works presented in chapter 3, which seemed to be dominated by three types of explanations. First, arguments of technical inadequacies in the models used to produce the forecasts, or in the available data material that are fed into them have been presented as possible explanations (Mackinder and Evans 1981; NAO 1988; Welde and Odeck 2011). This will typically be in the form of poor model calibration, limitations in model specification, or uncertain measurements for input variables. Second, arguments of cognitive biases have been offered as explanations, most typically in the form of optimism bias (Bain 2009; Fouracre, Allport, and Thomson 1990; Lovallo and Kahneman 2003). While typically a field explored in psychology and management, such theories are also supported by findings in neuroscience (see e.g. Sharot, Korn, and Dolan 2011). Third, arguments of deliberate manipulation have been offered, typically referred to as strategic misrepresentation. This terminology is borrowed from game theory where it has traditionally been used to describe strategies employed by negotiators to allow room for bargaining (Raiffa 1985), but its usage in planning literature has increasingly become synonymous with lying (Flyvbjerg 2005; Kain 1990; Pickrell 1990; Wachs 1989).

There are some limitations with the above categorization of explanations, of which the biggest is probably that they are not mutually exclusive. For example, a technical deficiency can be the result of both cognitive bias and strategic misrepresentation, and how to distinguish between them is not at all obvious. Furthermore, it is very difficult to provide credible explanations for these types of explanations in individual cases by simply observing the inaccuracy of forecasts. We must observe the process of how they came about in detail if we are to provide convincing analyses of this. This is why Kain's (1990) account of strategic misrepresentation in the DART/NCTCOG project remains so iconic (cf. subsection 3.2.5), since he had intimate details available to him that made such analysis possible. I have already described why this is an undesirable approach in the present study, but questionnaires and interviews might also be of some use here. Since it is practically impossible to identify the causes of inaccuracy in traffic forecasts with certainty (even Kain could only indicate that

this was a highly likely explanation for the observed outcomes) it does not seem unreasonable to aim for convincing indicators of tendencies among stakeholders. By asking them to describe the preparation, application, and scrutiny of forecasts that take place in the decision-making process, it should at least be possible to gain some pointers to which explanations are most likely to be of key concern.

As an extension of this issue we can also use the questionnaires and interviews to address obstacles in relation to the use of traffic forecasts in decision-making as they are identified by stakeholders. When faced with considerable uncertainty in the decision support with which they are presented, how do stakeholders believe this problem should be addressed? Do they merely accept that the future is uncertain, or do they call for increased clarity? Do they envision alternatives to model-based decision support? Such inquiry would provide inspiration for future research agendas in this field, but it would also be fruitful for the interpretation of the results related to all of the previous research questions. It might even call into question whether we need to adjust our assessment methods to address the needs of stakeholders more appropriately. The aim here is not so much to have stakeholders identify these obstacles explicitly, but to infer a set of problematic issues from the concepts that emerge when decision makers are asked to elaborate on the use of model-based decision support, and the way it is used to inform policy-making in relation to specific goals for transport planning.

4.4 Data sources

Having discussed the most suitable modes of inquiry for the present study we can now summarize them into the following four sources of data; impact assessments, databases, questionnaires and interviews. The assessments and databases will primarily be used for the first research question and make up a strictly quantitative approach. Questionnaires and interviews will primarily be used for the second research question, with the former being quasi-quantitative and the latter strictly qualitative.

The decision support has already been covered and shall only briefly be discussed again here. Ideally we would use the traffic forecasts that the EIAs and CBAs for individual projects were based on when selecting the reference point. However, these would be near impossible to recover for each project if we are to construct a sizeable dataset. Instead we shall use the available information in the impact assessments, with EIAs as the main source. Even these might be difficult to recover, so where suitable alternatives are available these will be used if the EIA cannot be recovered. This could include prior scoping reports if no changes have been made to the design and there is no reason to believe that assumptions of forecasting technique have changed. It is typical that the forecasts for these are reused in later impact assessments anyway in order to save resources for analysis. Another proxy could be the propositions made in parliament for individual projects, since these might also contain information from the impact assessments. However, this is usually in a very condensed form, and it can be difficult to specify which specific link a given traffic volume refers to in these propositions. Only when sufficient information is presented will these be suitable alternatives in lack of actual impact assessments. In addition to this it will not be possible to use these as a gauge for communication of uncertainty between stakeholders, so wherever possible we shall aim for EIA documents as sources.

Databases refer primarily to the count data that is available for road and rail traffic, which differ greatly between different periods, project types, and geographical location. Older counts for road traffic are mainly found in written reports and obtained via sporadic, manual counting techniques. This might also be true for less trafficked links in the case of newer projects. Permanent counting stations provide better validity for counts, but these might not be available for the desired reference point, or reports might be based on data for only a limited period of time (e.g. when ex-post counts are reported in connection with the opening of new schemes). Conversions to a representative value for AADT must then be made, but obviously these are themselves a source of uncertainty. For rail traffic the available count data is typically much less certain than for roads, as these have typically been performed manually, and continue to do so many places today. For example, many public transit institutions count the amount of passengers manually on one day every year, in order to trace the travel patterns of transit users. Obviously this is filled with a large range of uncertainties, and periodic counts are usually also conducted to monitor variations over time. However, these single day counts often remain the most comprehensive sources of data for ridership, and upon request will often be cited when prompting service providers for information on traffic volumes. This makes the counts sensitive to specific circumstances on the particular day of counting, such as weather conditions, service problems, or public gatherings that attract considerable traffic (e.g. concerts or sporting events). In order to obtain comparable data across the range of projects, traffic counts for road and rail projects will be based on the respective governing body in each country, while acknowledging the uncertainties involved in this approach. In addition to traffic counts are of course also databases over project costs, but these are generally poorly recorded and rarely available for direct inspection. Most ex-post cost data thus have to be retrieved via requests to the respective governing authority.

Questionnaires are sent to relevant stakeholders in the decision-making process for transport infrastructure planning in each country. Stakeholders are defined as the people that either help develop the models used to forecast traffic (e.g. modellers), apply them for individual projects (e.g. planners), base their decisions on their outcome (e.g. policy makers), or assess the validity of them (e.g. researchers). These descriptions are of course somewhat fuzzy as well, and individual stakeholders can easily take on a number of these roles. In addition it excludes a set of potentially relevant stakeholders, of which the most important is probably the public. However, including responses from laymen is drastically different from that of experts, and would likely necessitate a different set of questions entirely. While the respondents described above might have status of experts based on many different types of expertise, a common denominator for all must be that they are or have been used to dealing with traffic forecasts as a decision-making tool to some degree, while the general public cannot be considered to have much familiarity with the issue. Still, policy makers are likely less acquainted with many terms employed in technical aspects of transport planning, just as other stakeholders might have little insight into important aspects of policy-making. Certain questions on technical and policy aspects will thus be limited to the relevant respondents, while other questions will be general to all respondents. The questionnaire is conducted via an online form sent to the largest consultancy firms, public agencies, political groupings, and research institutes that work with transport planning in each country.

Interviews are made with key stakeholders to reveal more in-depth perspectives on certain issues. The definition of stakeholders remains the same as the one given above for questionnaires, and respondent selection is largely based on a snowballing approach. That is to say those respondents are selected from interview to interview, either due to reference from previous respondents or to highlight issues that came into light during the investigation in general. Respondents are not selected with particular concern for sample representativeness for a larger population, but rather to elucidate some of the points that have already seen broad representation through other investigations or to uncover new concepts that have so far been overlooked. However, representativeness in terms of their role in the decision-making process is of some concern, as we are obviously interested in the perspective of different types of stakeholders. The interviews themselves are open to semi-structured to allow respondents to address the issues they identify as most important, and structure is mainly confined to themes such as transport planning, forecasting, uncertainty, impact assessments, and decision-making. The interviews are anonymous to counter potential reservation in relation to the discussion of sensitive topics. Recordings and transcriptions are done of each interview and made available to selected members of the UNITE group, unless individual respondents object to this. All interviews are conducted with the author present.

4.5 Summary

The purpose of this chapter has been to outline the most useful modes of inquiry for the problem formulation defined in chapter 1. Clear guidelines were established for the selection of projects to include in the comparison of forecast and observed traffic and costs, although considerable difficulties of both practical and theoretical nature were identified. Decision support documents in the form of environmental impact assessments and cost benefit analyses were identified as the most suitable sources of forecasts, and databases from the respective governing authorities as the most suitable for counts. Questionnaire and interview responses were identified as the most suitable approaches for investigating uncertainty management among stakeholders.

5 Data collection

Errors using inadequate data are much less than those using no data at all.

- Charles Babbage

In this chapter I shall provide a brief overview of the empirical data that has been collected for the present study. Whereas the last chapter was chiefly concerned with elaborating on the rationale behind the chosen methods of analysis and the necessary data to conduct it, I shall here elaborate on how the data actually came about. The main purpose is to discuss some of the compromises that had to be made during the project, and how these affect the validity and reliability of the later analysis. As expected, data availability was somewhat lacking, and many approximations were necessary to establish reasonable datasets for both quantitative and qualitative data. I shall therefore also discuss a set of process related issues from this experience that will form a basis for part of the discussion in chapter 8. It also allows many of the more tedious descriptions to be grouped in a separate chapter than the analysis, so the latter might be presented without unnecessary cluttering from a range of conditionals for each finding being made therein.

5.1 Decision support documents

Decision support documents were described as the primary source of data for traffic forecasts in chapter 4, and are thus quite important for the validity of much of the analysis related to the first set of research questions. I argued in favour of using EIA and CBA documents available at project approval, to ensure that the analysis compares outturn traffic with the expected volumes used in the preparation of decision support. Denmark, Sweden and the United Kingdom are all members of the European Union, and are thus by law required to undertake environmental impact assessments of larger construction works. The projects of interest for the present study typically always fall into this category, as new transport infrastructure affects many aspects of the environment through construction as well as usage. Norway has also adopted similar requirements for construction works, and while the practice differs somewhat in the four countries, they are conceptually identical. The same can be said for the use of cost benefit analysis, which is the standard tool for appraisal of economic impacts in all four countries.

The approach to impact assessment changes periodically in most countries, and these four are no exceptions. For example, the United Kingdom has long been on the forefront of developing new appraisal techniques for impact assessment of transport projects. This led to the adaptation of the New Approach To Appraisal (NATA) in the late 1990s (DtF 1998), which sought to merge EIA and CBA results. While the term was officially abandoned in 2011, it remains the underlying methodology for impact assessment for transport schemes in the United Kingdom. However, common to all of these approaches is that traffic forecasts are used in the assessment of economic and environmental impacts, and has been so since the 1950s. It therefore seems reasonable to assume that traffic forecasts have been used in the preparation of decision support for most transport infrastructure

projects completed in the last couple of decades in all four countries. Techniques for modelling traffic have of course also changed over the years, but most have at least reported AADT values as a standard unit of measuring traffic volumes for roads and ridership expectations for rail. These measures have therefore been used as a basis for comparison of forecasting accuracy between traffic forecasts for projects from different countries and different time periods. For an evaluation of the impacts of traffic forecasts on subsequent impact assessments for road projects it is of course not enough to consider only these, but since detailed information on expected rush hour traffic and modal split is difficult to obtain for large samples, a comparison of expected and actual AADT or ridership values will have to serve as a proxy for forecasting inaccuracy.

In order to obtain the necessary decision support documents I initially started out by contacting the responsible directorates for road and rail planning in Denmark. The underlying rationality was that it would be easiest to start gathering data in my country of birth. Once a systematic approach to obtaining decision support documents had been established here, I had hoped to mirror this in the other three countries. However, it quickly became apparent that systematic storage of such information was nowhere to be found. Both the road and the rail directorate claimed that no such archive existed in their organizations, and after contacting the ministry of transport I was given the same message here. I thus requested a list of completed projects in the last 20 years, in order to gather information for each project individually, but once again the response was that no such list existed. At best I could use whatever available online resources to piece such a list together myself, and I therefore began constructing a list with the aid of a few helpful individuals.

Whether the lack of information was due to the contacts within the directorates being unable or unwilling to help was of little concern to me at this point. It might have been the case that these people simply did not know where to find this information, but after unsuccessful requests to receptionist, librarians, project managers, and administrative coordinators alike, I felt that if such archives did in fact exist, they are at least not made accessible for research. Instead, a list of projects was circulated in the UNITE group, and with the help of contacts in both the road and rail directorates I ended up with a list of roughly a hundred completed projects. I then returned to the road and rail directorates for assistance in locating the respective decision support documents, but once again my request was unsuccessful, and the response was that I would have to obtain these documents by locating them myself or supply a list of named reports and publication years.

It seemed fairly clear that if any options existed for tracking this information internally, they were at least not known by any of the people I contacted. I should like to note that at this point I was becoming fairly sceptical about the transport planning process in Denmark since my discoveries so far had two possible interpretations. If the main problem was that these people were *unwilling* to help, then open access policy to public documents was severely hampered by obstruction from public officials refusing to help provide such information. If the main problem was that these people were *unable* to help, then this is a clear indication that no systematic ex-post appraisals of transport infrastructure projects are carried out, and thus that very limited attention is given to how well projects fulfil their objectives. None of the interpretations are particularly comforting for a researcher in the field of transport planning, but since most of the people I have been in contact with seemed genuinely interested in both the research topic and in helping me progress, I would have to

conclude that the latter option is probably the most realistic. Regardless of the cause, no structured archived was available to me. If the issue of data availability was not made clear in chapter 4, I hope the reflections presented here have helped enforce this message, and as the chapter progresses it shall hopefully become even clearer how constricting this issue has been for the present study.

Charged with a list of completed projects I thus began scouring online resources for available decision support documents. Many of the most recent publications were available via these sources, but going back further than ten years quickly proved problematic. As even smaller transport infrastructure projects typically have a timespan of 5 years or longer from political decision to project completion, most of the projects on my list could not be covered by this approach, since the decision support documents for them were often prepared around 10 to 30 years ago. I thus began searching elsewhere for these documents, including libraries, public archives, and by personal request to people that had worked on the respective projects. It should be obvious that this approach has been fairly time-consuming, and a great deal of resources has been devoted to this task throughout the project. It is certainly possible that data was readily available for many more projects if I had known the right places to look, but if nothing else the approach has probably ensured that the sample I managed to construct is as randomly selected as possible for this type of study. Rail projects proved particularly problematic to uncover documentation for, both because there are a fewer of them and because demand forecasts were not always included in the available decision support. In the end, it was possible to construct a decently sized sample of Danish projects, but it was highly resource demanding to do so.

The experience was sadly not much different in Norway and Sweden, where the road and rail directorates proved unable to assist with the acquisition of archived decision support documents. Furthermore, where each project in Denmark is passed through an associated construction law, both Norway and Sweden have national transport plans that are updated every four years instead. This made it somewhat problematic to identify the reference points for when a project actually gained political approval in its final form. A research assistant was temporarily employed in the project to help obtain additional decision support documents at this point, and while this helped speed up the process slightly it did not overcome the obstacle of availability. In addition, it quickly became very resource demanding to gather data through physical archives in three different countries. Given that the UNITE project had a partner in Norway and none in Sweden, I decided to employ the Swedish EIA Centre in Stockholm to gather data for a smaller sample of projects, hoping that they would be able to access this information more easily due to experience and local knowledge. Meanwhile, I could focus on data collection in Norway with the assistance of our UNITE partner at the Institute of Transport Economics in Oslo. However, the Swedish EIA Centre faced similar problems in obtaining the necessary documentation, and was able to supply only scattered information for a limited number of projects, and with relatively poor documentation of sources as well. At this point I decided to abandon further inquiries in Sweden, as it seemed undesirably resource demanding to increase the sample. In Norway it was possible to gather a great deal of EIA reports in a short period of time, since the Norwegian Institute for Urban and Regional Research (NIBR) had begun archiving EIA reports for all major projects in Norway sometime during the 1990s. While this initially gave rise to some optimism regarding the possibility for a large sample of Norwegian projects, it turned out that many of these projects were either not yet completed, had later been amended by changes made

prior to political approval, or did not contain the necessary data for the analysis in the present study. The sample for Norwegian projects thus also remains fairly modest.

In the United Kingdom it was possible to construct a reasonable sample for road projects in a relatively short period of time, since the Highways Agency has conducted the Post Opening Project Evaluation (POPE) programme for some time now. This involves assessments of major schemes one and five years after opening, and in these assessment reports there is a great deal of information available from the original decision support documents. While it would have been preferable to use the actual EIA and CBA documents as sources, the POPE reports seemed a suitable alternative, since they were based on data from the original decision support for individual projects. Additionally, I assume the Highway Agency has much better access to these sources than I could hope to achieve. Given that the necessary data collection efforts in the other three countries had already been considerably greater than anticipated, I decided that this was justifiable compromise to the methodology outlined in chapter 4 in order to reach a useful sample size. Going through these reports it also became clear that even though the Highways Agency had commissioned an internal review of completed projects, they seemingly had similar difficulties in obtaining the forecasts used in the original decision support documents. While not of particular comfort to the goals of the present study, it was somewhat reassuring to experience that access to data was indeed a concern internally in the directorates, and seemingly not because officials were withholding such information deliberately.

In the end it was possible to obtain traffic forecasts for roughly 250 projects in the four countries combined. However, not all of these were based on the original EIA or CBA documents for the projects. In some cases it had only been possible to obtain strategic analyses that contain impact assessments for a given project, but where it is highly likely that additional analyses has been carried out prior to political approval of the project. However, in many EIA documents it is explicitly stated that the forecasts are simply based on such prior analyses, and it does therefore not seem unreasonable to assume that these are reasonable proxies in many cases, if no significant changes in design has since been made. For other projects it has been entirely impossible to obtain any of the actual decision support documentation, but it has instead been possible to find information on expected traffic volumes in the proposed construction laws. In these cases I decided to use the construction law proposals as sources if the information is described in sufficient detail to identify the particular location that the expected traffic volumes refer to. At any rate, this is only the case for a small part of the projects in the final sample.

5.2 Databases

As was the case with decision support documents, databases form an important source of data for the first set of research questions. Generally speaking it has been much easier to access reliable data for traffic counts than it has been to access decision support documents for a number of reasons. First, traffic counts are tracked much more systematically than are forecasts, as there might be many different forecasts prepared prior to the opening of a project. Counts are uniquely identifiable to a specific spatial and temporal reference point, making it easy to construct a database for this. Conversely, forecasts might correspond to projects that never materialize or appear in multiple versions at a given time, and are thus not uniquely identifiable to a specific spatial and temporal

reference point. Second, few people appear to have any interest in the traffic forecasts after they have been used to prepare impact assessments, while traffic counts are quite important for continued monitoring of the system network. There is thus no real incentive for stakeholders to archive forecasts systematically after a political decision has been taken. Third, since forecasts are more difficult to access than counts, availability of forecasts became the initial selection criterion for projects to include in the database. There was thus a smaller pool of projects for which to obtain counts than there were for forecasts, since the potential pool of projects had already been reduced in size.

However, this does not mean that obtaining traffic counts have been devoid of problematic issues. The first problem that presented itself was the representativeness of the counts to reflect general traffic levels. For example, counts might not be monitored permanently on a given road stretch, meaning that we have to infer AADT values based on counts at various points in time over the year. There are standard statistical procedures for such scaling procedures in the road directorate of each country, but this obviously adds some uncertainty in regards to how accurate these approximations are. Obviously the reliability of the counts increase with longer periods of observation, and as a rule of thumb I have tried to avoid using counts based on anything less than four weeks of observation over a year for road projects. Since data for the projects from Sweden and the United Kingdom has been based on results from the UK Highways Agency and the Swedish EIA Centre, I thus had limited control over the reliability of the counts used for these projects. However, it seems reasonable to assume that the respective organizations are both perfectly aware of such issues, and that quality checks are in place for the reporting of counts based on shorter time periods. For Denmark and Norway I was granted access to the official count databases from the respective road directorates, while counts for rail are based on annual statistics or database extractions from the respective rail directorates or service providers. The common use of a single day of counting each year for rail projects has already been mentioned in chapter 4, and there are obviously a number of uncertainties in regard to representativeness stemming from such an approach.

Another problem presents itself in comparability between forecasts and counts, since these might not be based on the same units or they have different spatial or temporal reference points. For example, one project might report rush hour traffic in the forecast but provide no definition of rush hour. In this case I assume that rush hour refers to the busiest hour span of the day, which is typically 15:00-16:00 o'clock, and while it might seem a reasonable assumption it is nonetheless associated with some uncertainty. It might also be the case that no counts are available for rush hour traffic, but only for the daily average. In this case we might assume that rush hour traffic makes up a certain percentage of daily traffic volumes, but this is also an assumption filled with some uncertainty. For rail projects the problems are typically much greater, since it is obviously very difficult to infer the travel patterns accurately from a single day of counting. On top of this the forecast might only describe ridership between two destinations, and since counts are typically done station wise how do we infer the number of people traveling between two destinations when we only know how many people have gotten on and off at these points? Even more problematic are situations where ridership forecasts refer to a single transit line in an urban area where multiple lines might overlap. How do we then discern between traffic on a newly constructed line and three existing lines if we only have counts available at each end point, if the forecast only reports the amount of ridership that the new

line will attract? Unless these issues have been specifically accounted for in the questionnaires used for the single counting days there are not really any good answers to these questions, and so counts for rail projects are crude approximations at best if there are inconsistencies between the unit of traffic used in forecasts and counts. When I have approached the rail directorates for help with these issues they have typically been quite reluctant to offer any specific method of conversion, claiming that it would largely be guesswork to do so. Apart from presenting a set of problematic issues for comparison of forecasts and counts, it is also further indication that no systematic ex-post appraisals are carried out to investigate whether ridership expectations have been met.

Even if counts for road projects were much better than for rail projects, important compromises still had to be made. For the vast majority of projects, forecasts were presented for an entire system network, including existing roads in the same corridor. However, counts were often only available for a few of these locations, and the comparison of reference point in the present study is therefore limited by this inconsistency. Sometimes there were no overlapping reference points for forecasts and counts at all, and in such cases I either scrapped the project entirely or made do with interpolation of traffic volumes from nearby reference points. The latter option obviously requires that no important points of attraction or generation for traffic between these reference points exists, such as a rural setting with no urban settlements or work places in-between. Even so, this is a subjective call for each such case, and naturally associated with some uncertainty. In addition to all these sources of uncertainty are of course also errors or fluctuations in traffic, which are not necessarily insignificant. For example, for some projects traffic volumes reported in count databases would differ as much as 100% between two subsequent years. When contacting the road directorates about such cases, no plausible explanation could usually be given for these differences, and on several occasions I have been thanked by them for pointing out oddities or mistakes in the counts. While always a pleasure to be of service to other people, these findings did not really bolster the present study, as it was not always possible to find a good approach to dealing with these inconsistencies. On top of this are of course general uncertainty related to the data material, since large inaccuracies could potentially be caused by errors such as those reported above, in case they go unnoticed. Ideally, it would be desirable to employ observations for a span of several years to address these issues, but due to aforementioned problems of data availability this has not been possible in the present study.

Data for links other than the main project links have been particularly difficult to obtain for many projects, since these are typically managed by local authorities, which in many cases have no permanent counting stations. This results in poor availability of data, and even if relevant data exists, it is typically less reliable than for the major links. Where reliable counts have been available, these have been included in the analyses, but many projects only have available data for the main project links. Similarly, comparable traffic count data for the zero-alternatives have been very sparse, since this requires at least one of the following three conditions to be fulfilled. Option one is that the project has been abandoned. In this case it is a true zero-alternative, where observed development can be compared with the expected development. Option two is that construction starts several years after the forecast is made. In this case the growth trends between when the forecast was made to when construction started can be used to verify expected growth trends for the zero-alternative. The third option is that construction does not affect traffic on the existing network. In this case the

growth trends from when the forecast was made to when the project opens can be used to verify expected growth trends for the zero-alternative. A final comment on traffic counts should be made for projects that opened more than one or two decades ago, since these are typically not found in digital form. Instead I have had to rely on annual count statistics published by the respective directorates where these have been available. Many of these are manual counts and thus associated with greater uncertainty than many of the permanent counting stations that are now standard on larger links. Apart from this I doubt there are any particular problems related to using these as a source of data that have not already been covered.

Regarding costs, this has been much more problematic to get reliable data for. There are three main reasons for this. First, both initial budgets and final costs are often scarcely documented in the available decision support and ex-post reporting, making it difficult to identify whether the items in the budget and the final expense report are identical. Budgets can obviously vary drastically depend on whether or not they include administration, design, or land acquisition. If these are explicitly excluded in the budget but included in the figures for the final costs it is no wonder that the project will appear to suffer from cost overruns. This was the cause of much controversy regarding the expansion of a Danish rail project³⁰, where an early draft covering only raw construction costs of new rail facilities was used as decision support. The resulting accusations of the incompetence of cost estimates from the Danish State Railways (DSB) caused many of the responsible engineers to feel as scapegoats, since they had never claimed the estimate to be an assessment of the total project costs (Marfelt 2010; UNITE 2011b). During a personal visit to the department responsible for cost appraisal in Swedish transport planning, several people pointed out that they also faced this problem internally when trying to make internal ex-post project evaluations. A new motorway project might appear more expensive than expected at first sight, but on closer inspection it might turn out that this is because a bridge for a perpendicular link crossing the new motorway project was included in the final expense report, but not in the initial budget. However, at that time it might have been part of a budget for another project. An ex-post evaluation based on such figures would conclude that the project faced cost overruns, whereas the funding had already been granted through a different budget.

Second, many of the projects relevant to the present study were completed some time ago, and systematic records of cost data do not appear to be very strong in any of the case countries. Even relatively new projects that were completed less than a decade ago have been very problematic to uncover reliable cost data for. If any useful information exists at all, it is typically in aggregate figures that are difficult to compare with specific budget items. Some of the projects have had their budgets approved as a political agreement on a joint project package, but where it is only possible to access the final expense figure for the joint project package. It is obviously quite difficult to assess the cost performance of individual projects in the package, when either budget or final cost figures (or both) are only available at an aggregate level. Adding to this confusion has been the institutional restructuring of the responsible government agencies that has taken place in many of the case countries. In both Norway and Sweden the road and rail directorates have recently been merged, and information requests are often stopped dead in their tracks simply due to the difficulty of tracing

³⁰ The double track expansion of Frederikssundbanen that opened in 2002.

documents prepared prior to this merger. In Denmark the respective directorates are still separated, but the rail sector has seen a similar split of institutional responsibilities between planning and service agencies, with equal confusion regarding storage of documents prepared under different institutional settings³¹.

Third, even if a more detailed breakdown of costs exists, this does not mean it is publically available or accessible for research purposes. It has been my clear experience that while requests for passenger data for rail services might result in raised eyebrows and sceptical attitudes from the responsible authorities, request for cost data is almost certain to be met with grave concerns about the nature of such inquiry. For some projects I was able to find ex-post evaluation of cost data in reports from the respective audits of state accounts, but when attempting to retrieve sources or specific details for the data used in these evaluations I was usually directed from one person to another between various public authorities, until finally informed that the information was no longer accessible or that it was confidential. Regardless of policies to ensure freedom of information in public affairs, it is seemingly quite difficult to get much detailed information about how public money is being spent. From a practical perspective it quickly becomes unfeasible to spend several weeks of communication with public authorities to obtain the necessary documentation for a single project.

5.3 Questionnaires

While decision support documents and databases are mainly related to the first set of research questions, the questionnaires are mainly related to the second set of research questions. As was described in chapter 4, the purpose of them is mainly to extract information about stakeholder opinions on a set of issues related to uncertainty in using traffic forecasts as a decision support tool for policy-making in transport planning. The idea was to obtain a set of quantitative measures for how the concepts of uncertainty and inaccuracy are perceived by different types of stakeholders, and respondents are thus asked to supply information about their experience with transport models and traffic forecasts, their educational background, what types of projects they are typically working with, what role they usually take in the decision-making process, etc.

Apart from general classification categories and a few generic questions on the perceived accuracy of traffic forecasts, the respondents were asked to rate their agreement with four sets of different types of statements. The first set relates mainly to the general perception of traffic forecasts and their usage as decision support. The second set relates mainly to specific model related issues. The third set relates mainly to potential causes of inaccuracy when comparing expected and outturn traffic volumes. The fourth set relates mainly to the use of forecasts in policy-making. Policy makers are not presented with the third set of questions, as these mainly concern technical issues for planners or modellers. Following a similar line of reasoning, only policy makers are presented with the fourth set. All statements in each of the four sets are ranked with a five point Likert-scale response format with the value 1 representing full disagreement and the value 5 representing full agreement. Likert-scales are commonly employed when comparing attitude responses, and while it is ultimately an ordinal scale I shall often treat it as an interval scale. There is considerable controversy

³¹ As an example, requests for specific cost details on Danish railway projects were often met with the same response from the rail directorate, ministry of transport, and audit of the state accounts alike: *“we don’t have accessible records of that information anymore”*

in the social sciences surrounding the proper classification of Likert-scale items and whether it is justified to consider response values interval instead of ordinal (Jamieson 2004). The controversy mainly relates to appropriate measures of statistical significance, and it is thus not without concern for these issues that I choose to treat questionnaire responses as interval data, suitable for tests of significance. I shall also employ non-parametric tests of results in cases where a relationship is not obvious, in order to ensure that any statistically significant relationship is not merely the result of treating responses on the wrong level of measurement. However, there are many other types of uncertainty associated with this type of survey method that pose greater problems than choosing the correct level of measurement, and often it will make only negligible difference to the interpretation of Likert-scale items. This is especially true for the present study, as the main purpose of the questionnaire is to get a set of indicators for relationships rather than verify a statistically significant relationship between a set of variables. Some scholars even argue against the use of statistical significance when reporting findings, since the measure is of no practical value, and often interpreted incorrectly even in leading scientific journals (Armstrong 2007). I have thus mainly focused on reporting visual indicators of results, as recent evidence suggests this to be a more effective way of communicating uncertainty in empirical studies (Soyer and Hogarth 2012). For the same reasons I have not spent a great deal of time considering whether a five, seven or ten point scale was the most appropriate measure.

Stakeholders were defined earlier as the people that either help develop the models used to forecast traffic (e.g. modellers), apply them for individual projects (e.g. planners), base their decisions on their outcome (e.g. policy makers), or assess the validity of them (e.g. researchers). However, it is obviously difficult to use these crude definitions to form an exhaustive list of possible stakeholders for each country, and no attempt at any such list have been made either. Instead, I have identified what I consider to be relevant communities of experts for each country, where a larger group of people who match these profiles are likely to be found. These include transport related research units at universities, transport units in consultancy firms, road and rail directorates, transport sections in regions/municipalities, transport/environment related NGOs, and parliamentary transport committees. In each case we asked for response from people who indicated that they had prior experience with traffic forecasts as decision support, either through constructing transport models, preparing traffic forecasting, forming transport policies, or critically assessing models/forecasts.

Without any exhaustive overview of the reference population it is practically impossible to account for whether this approach has yielded a representative sample, and I remain uncertain as to how in fact to establish such an overview in a meaningful way. Nonetheless, the institutions to which the questionnaire was sent must be considered among the key stakeholders in transport planning for their respective countries. The questionnaire was handled exclusively online, and links were sent to contact persons within each identified community in Denmark, Norway and Sweden, which yielded 453 individual responses. The United Kingdom was not included in this survey due to communication problems internally in the UNITE project, since I was informed of another questionnaire prepared for the United Kingdom. I thus postponed the collection of survey data from this country until I was sure that potential respondents here would not receive two different surveys from the same research project attempting to gather similar types of data. However, over a full year after the surveys had

been carried out in Scandinavia I still had no confirmation on the status of this parallel questionnaire, and decided to abandon the collection of data for the United Kingdom.

Due to overlapping research interests, the questionnaire was prepared in collaboration with Work Package 2 (WP2) in the UNITE project, in order to avoid sending out two different questionnaires to a set of informants that would largely be identical. Thus, not all items in the questionnaire are prepared explicitly for the present study, although they are of course still of some interest to it.

5.4 Interviews

As was the case for the questionnaires the interviews are mainly related to the second set of research questions, and since many of the issues related to the data collection from the questionnaires apply here as well I shall only briefly cover the most important additions here. Stakeholders were defined similarly to the target group of the questionnaires, but with a snowballing approach to selection of informants. As with the questionnaires there were many overlapping points of interest between the present study and other parts of the UNITE project, resulting in some of the interviews being conducted in association with Jeppe Astrup Andersen from UNITE WP2. Based on similar reasoning as that given for questionnaires, interviewees from the United Kingdom have been omitted from the interviews. The total number of interviewees is 14, of which the majority is from Denmark. As was discussed in section 5.1 there were considerable resource constraints associated with gathering useful decision support documents in the rest of Scandinavia, and I thus chose to focus on including informants from the country where the best reference material was available.

As was mentioned in chapter 4, the purpose of the interviews is not to achieve a sample of informants based on their representativeness in % of total stakeholders. Given that no overview of the reference population can be made, this would indeed also be impossible. Rather, the purpose is to achieve a sample of informants that represent different types of stakeholder perspectives. Informants included are thus selected based on two principal criteria. First, they must represent a key stakeholder in the decision-making process according to one of the four categories defined in section 4.4. Informants thus cover a wide range of different stakeholders, including consultants, researchers, public officials, and parliamentary politicians. Second, they must have substantial experience with transport planning in their respective country, either from strategic planning or specific projects. Substantial in this relation taken to mean at least 5 years of working experience within the respective position, but typically the selected informants have been working with transport planning for a considerably longer period. Informants with the longest work experience in the field are typically project managers in consultancy firms or public administration, while those with the shortest work experience are typically politicians. However, in the events of relatively short working experiences in the field of transport planning, the informants have typically been chosen because of their familiarity with a particular issue of interest, and are thus considered experts within the scope of the interview. While this was not something I considered particularly during the selection of informants, it does not seem unreasonable for this to be a general trend outside the sample informants. Project managers for larger infrastructure projects are typically required to have significant working experience in the field, while parliamentary politicians do not necessarily build up

expertise in the field over extended periods of time for a variety of different reasons (e.g. because they are not re-elected or move on to responsibilities other than transport planning³²).

5.5 Summary

The purpose of this chapter has been to provide an overview of some of the main obstacles encountered during the collection of the necessary data items outlined in chapter 4. There were no initial plans to include such a chapter, but the considerable difficulties in obtaining reliable ex-ante and ex-post data have themselves proved important observations. Systematic archiving practice for decision support documents was found to be rare in all case countries, and access to the necessary count data has also been restricted in various ways. This led to a substantial amount of resources being spent on obtaining the necessary data for the first part of the analysis (chapter 6), and consequently a reduced amount of resources available for the second part of the analysis (chapter 7). The difficulties described here are highly relevant in the interpretation of the observed forecasting inaccuracy, and likely pertain to other studies of forecasting accuracy as well.

³² As an illustrative example, Denmark had five different ministers of transport from 2007 to 2010, of which most either moved on to other policy areas or out of politics entirely upon leaving office. This might be because transport policy is unlikely to be an area that attracts much voter support. Even if projects turn out to be successful upon completion, the responsible decision makers are often no longer around to reap the benefits of potential increases in voter support. Meanwhile, complaints regarding any contemporary problems such as road congestion or lack of transit reliability are almost certainly directed at the people currently in office. This reasoning obviously holds true for many policy areas, but especially so for transport planning due to the long time periods from planning to completion of projects. A particularly unpopular aspect is the potential expropriation of residents in relation to acquisition of land in the early phases of a project.

6 Forecasting inaccuracy

It is easy to lie with statistics, but easier to lie without them.

- Frederick Mosteller

In this chapter I will present what I consider the most important findings from the comparison of forecasts and observations. This naturally begins with a thorough description of the forecasting accuracy for traffic demand for both road and rail projects, as well as a short description of forecasting accuracy for construction costs. The inaccuracy distributions for these measures form the backbone of the analysis, and will be used as reference points in the rest of the chapter. After discussing the observed inaccuracy in detail I investigate whether there are any specific project characteristics that are associated with different levels of forecasting accuracy. Some of these are possible to test statistically, while others require more qualitative assessments. Wherever possible I shall attempt to devote equal attention to road and rail projects for a given item in the analysis, but for certain items it might be necessary to elaborate more extensively on characteristics for one group while the other is largely ignored, either due to lack of relevance or data availability.

6.1 Forecasting accuracy

The first research question (cf. section 1.3) largely relates to the extent of inaccuracy in traffic forecasts, for which forecasts and counts were collected for statistical analyses. The total project sample for road projects is 146, of which the majority is located in Denmark and the United Kingdom. The total project sample for rail projects is 31. A more detailed breakdown of projects for each country can be seen in Table 2, where data availability for different types of comparisons is also shown. Project schemes are the main links of individual projects, whether they are entirely new links or upgrades to existing links. Existing networks are the reference networks that are not physically affected by the new schemes, but where traffic flows are expected to be influenced by them. Zero alternatives are the reference scenarios in which the project schemes are not implemented. Station breakdown are projects where both counts and forecasts are available for specific train/metro stations.

As can be seen in Table 2, data availability has been highest for the main project schemes in all countries when compared to data for the existing networks, zero alternatives, and station breakdown. This is fairly unsurprising, since ex-post traffic flows on the new links are typically of most interest for road projects and thus better monitored than the surrounding network. In addition, the existing network often consists of smaller roads, where counting might only take place periodically and for specific purposes. In these cases there is no data available for comparison, unless ex-post appraisals of the projects are made immediately after opening and the existing networks are considered for such appraisals. This is typically the case in the United Kingdom due to the Highways Agency's POPE initiative. Consequently, most of the projects in this category are located there. A similar argument can be made for the zero alternatives, where the ex-post appraisals in the United

Type		Denmark	Norway	Sweden	United Kingdom	Total
ROAD	Project scheme	75	23	10	38	146
	Existing network	22	0	1	32	55
	Zero alternative	15	0	0	20	35
RAIL	Project scheme	9	6	11	5	31
	Station breakdown	4	1	0	0	5

Table 2: Overview of road and rail projects in each country, for which comparable forecasts and counts have been obtained for different types of comparisons. Countries with a relatively high number of projects of a specific type is either indicative of substantial data collection efforts (e.g. road projects in Denmark) or above average data availability (e.g. road projects in the United Kingdom).

Kingdom have served as a good source of data for pre-opening counts. The reason why projects from Denmark are also relatively well represented in these two categories is entirely due to increased data collection efforts compared to the other countries. It might have been possible to obtain similar data for a set of Norwegian and Swedish road projects, but it was not possible to get hold of this within the time limits of the present study. For rail projects, most of these only have data available that allows for a rough comparison of expected and actual passenger volumes on the overall service line. Where station breakdown is available, it is typically for projects located in denser urban areas. It has generally been very difficult to find useful data for rail projects, and a large number of projects were dismissed because of insufficient or unreliable data. In the remainder of the present thesis I shall refer exclusively to data for the project schemes when I discuss the accuracy of traffic forecasts, unless otherwise stated. I have compiled all data on completed projects in a central database to make it available for the rest of the UNITE research project, and I refer to this database for all sources of individual data items (UNITE 2011a)

6.1.1 Forecasting accuracy for road projects

To get an idea of the scope of inaccuracy we can begin by looking at the distribution of inaccuracy for road projects, which is shown in Figure 20. As discussed in chapter 3 we would expect the distribution to be approximately normal, but possibly with a heavy tail in the positive end of the inaccuracy spectrum³³. This description matches the distribution of inaccuracies in Figure 20 rather well, as we observe a concentration of projects around zero, but with a heavy right-sided tail and a non-zero mean. Around two thirds of the projects have observed traffic volumes that fall within $\pm 20\%$ of the forecasts. Roughly 10% of the projects have attracted less than 80% of the forecast traffic volumes, another 13% of projects have between 20-60% more traffic than forecast, and the

³³ The intuitive explanation for this is that our inaccuracy scale has a theoretical minimum but no theoretical maximum. In practice there are of course upper limits to how much traffic can rise above forecasts due to capacity constraints, but these might be very large for certain projects. Some of the existing network links in the present study have an observed inaccuracy of more than 1200% due to forecasts being much lower than available capacity, but these are usually of little importance and therefore excluded from the analyses.

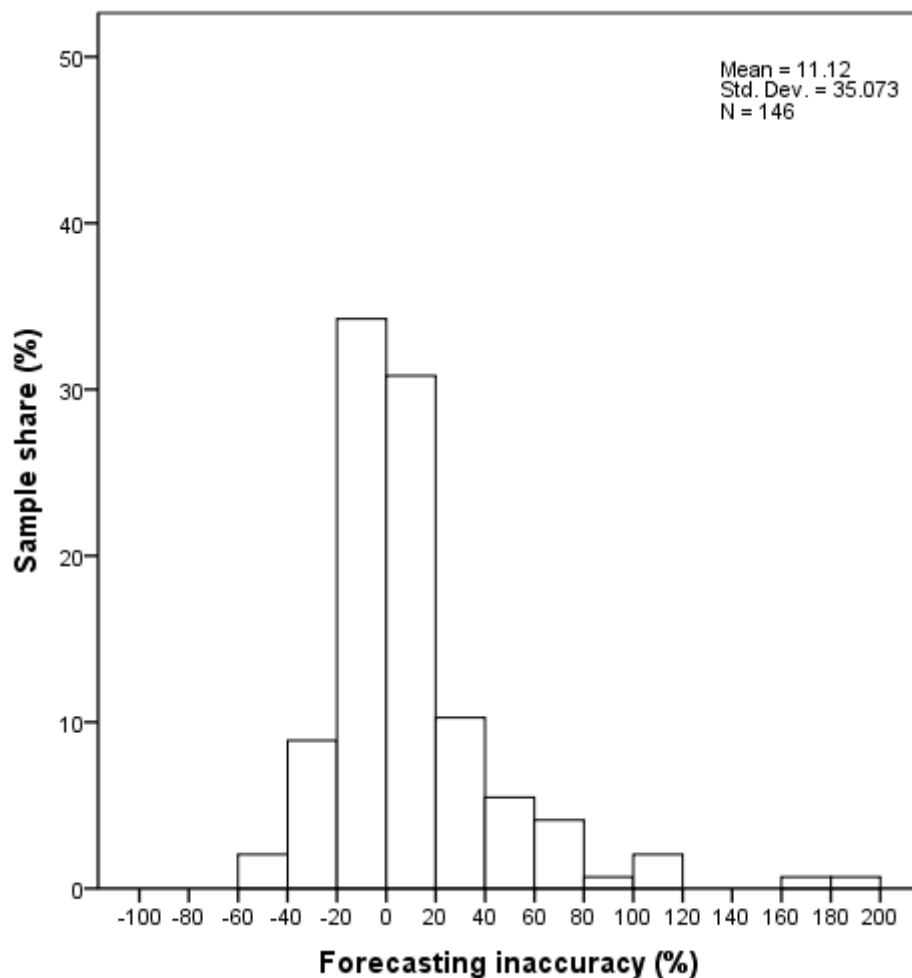


Figure 20: Inaccuracy of traffic forecasts for road projects in the total sample.

last 10% of projects has over 60% more traffic than forecast, with a few projects having almost three times as much traffic as forecast.

Table 3 displays the most important descriptive statistics of forecasting uncertainty for road projects to gain a better understanding of the observed distribution in Figure 20. We see that in spite of the large spread in the sample, the mean is significantly different from zero. Since the previous studies presented in chapter 3 have very similar means and standard deviations, it is not unreasonable to assume that the true population mean is closer to the sample mean than the confidence interval in Table 3 implies. The truncated mean is of course slightly smaller due to a heavy right-sided tail, but still indicates an average underestimation of traffic in forecasts. Positive skewness and kurtosis is further evidence that forecasting inaccuracy is not entirely normally distributed. A median of 1.64 indicates that slightly more than half the projects experience more traffic than forecast, and is further evidence of a skewed distribution since it is lower than the mean. Overall, the additional information in Table 3 indicate that traffic forecasts are only slightly more prone to overestimate traffic than to underestimate it, but that overestimations generally deviate further from the mean than underestimations.

Inaccuracy		Statistic	Std. Error
Mean		11.12	2.903
95% Confidence Interval for Mean	Lower Bound	5.39	
	Upper Bound	16.86	
5% Trimmed Mean		7.75	
Median		1.64	
Variance		1230.146	
Std. Deviation		35.073	
Minimum		-45	
Maximum		194	
Range		239	
Interquartile Range		26	
Skewness		2.265	.201
Kurtosis		8.019	.399

Table 3: Descriptive statistics of forecasting inaccuracy for road projects.

If we compare the observed distribution with the previous studies of road projects presented in chapter 3, we find that it is pretty similar to the majority of these. Table 4 lists the sample size, mean, and standard deviation for the five previous studies of non-toll road projects to compare with those of the present study. The findings in the present study are much in line with three of the previous studies, and any differences are more likely to result from standard errors of the mean than any differences in the reference populations. The findings are less in line with the two smaller studies. For Mackinder and Evans (1981) this is likely a result of assumption drag, where growth trends from a period of economic growth was used to predict development in a period of economic recession. For Welde and Odeck (2011) the sample is small enough that a confidence interval of the mean would most certainly overlap with the larger studies, making it difficult to assess whether forecasting accuracy actually differs between the samples. It is also possible that Norwegian projects have a tendency to severe underestimating of traffic due to ignorance of induced demand, as Welde and Odeck (2011) themselves point out as a plausible explanation. I have excluded the toll projects from Bain (2009) and Welde and Odeck (2011) from this comparison since they appear to behave quite different to non-toll projects. A small number of toll projects were available in the sample of the present study, but they do not deviate in any statistically significant way from the rest of the sample. This is not to say that the findings of Bain (2009) and Welde and Odeck (2011) are not interesting to compare with those of the present study, but the lack of project specific details in these studies makes it impossible to assess whether this is solely a result of them being toll financed. I have therefore chosen to keep the few toll projects in the total sample of the present study, as the results are unaffected by their inclusion. With these reservations in mind it seems fair to use a mean of 10 and a standard deviation of 40 as crude indicators for the distribution of forecasting accuracy for road projects, as indicated by the results presented in chapter 3.

When comparing the findings in Figure 20 with those of previous studies from chapter 3, we also observe that the distribution in the present study appears to be centred more tightly on zero. This

Author(s)	N	Mean	Std. dev.
Mackinder and Evans (1981)	44	-7	N/A
National Audit Office (1988)	161	8	43
Flyvbjerg et al. (2005)	183	10	44
Parthasarathi and Levinson (2010)	108	6	41
Welde and Odeck (2011)	25	19	21
Nicolaisen (2012)	146	11	35

Table 4: Comparison of forecasting inaccuracy for road projects between the present study and the largest previous studies where comparable data exists.

should be obvious merely through visual comparison with the histograms presented throughout chapter 3. Two explanations for this phenomenon immediately spring to mind. The first is that the complexity of the projects might be small in the present study, which would certainly make it easier to forecast demand more accurately. This hypothesis is difficult to validate, but there is no immediate evidence that suggests this to be the case. The sample in the present study is mainly comprised of state roads, with the majority being highways. This appears to be the case for the other studies as well, and while the sample from Flyvbjerg et al. (2005) is often associated with mega projects (e.g. Flyvbjerg, Bruzelius, and Rothengatter 2003; Flyvbjerg 2007) there seems to be no clearly defined boundary for what constitutes a mega project. The span of their sample includes projects with a cost of just \$1.5 million (1995 prices), which is lower than any of the projects in the present sample. With no additional information on the cost span it is difficult to gauge whether the samples are comparable in this regard, but if total cost can be used as a rough indicator of complexity there seems to be no evidence in favour of the first hypothesis.

The second hypothesis is that the adjustments I have applied for mismatching reference points have reduced the general inaccuracy measures. Since the main reason for applying these adjustments is to avoid an inflated measure of inaccuracy for individual projects, it should not be a controversial assumption that the unadjusted inaccuracy figures have a larger spread than the adjusted figures. Unlike the first hypothesis there is at least evidence that some of the previous studies have not made any such adjustments, and if this has a large effect on measured inaccuracy this is in itself an important finding. The available data for the previous studies does not allow such a comparison, but we can at least compare them for the sample in the present study. Figure 21 displays the distribution of forecasting inaccuracy for road projects using the unadjusted forecasts and counts, and it is clear that this provides a very different picture than using the adjusted values. As expected, the spread is slightly larger when no adjustments are made, just as fewer projects are centred on zero. However, the most striking feature is the drastic increase of the mean, since the majority of adjustments have

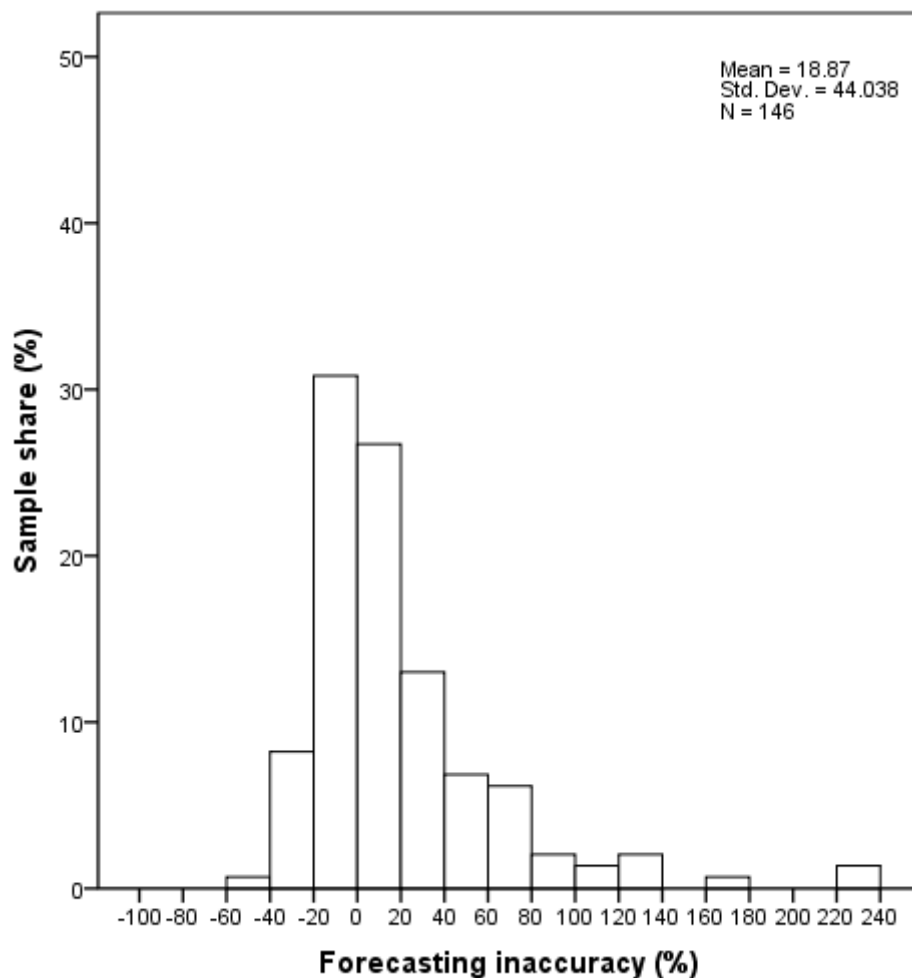


Figure 21: Inaccuracy of traffic forecasts for road projects in the total sample, unadjusted for mismatching reference points between forecasts and counts.

been for projects that opened later than the forecast target year. Some projects are actually not included in this figure, since the unadjusted inaccuracy exceeds 200% by a large margin. However, for the sake of comparison with other figures I have decided to use the same scale boundaries.

It is important to note that the findings on adjusted vs. unadjusted figures should not be taken as an indication that road projects are typically delayed, since the forecast target year is not always an expression of expected opening year. In fact, this mismatch might even be intentional, and the adjusted figures thus offer a much better measure of inaccuracy than the unadjusted figures. If projects are typically delayed this is certainly an important issue in itself, but not one that should be assessed by comparing forecasts and counts for mismatching reference points, since there could be other reasons for such deviations. If one wishes to undertake such analysis, it would be much more straightforward to simply compare expected and actual opening years for projects. At any rate, the large differences between adjusted and unadjusted figures reported here should be kept in mind, whether one decides to use one or the other as a measure of inaccuracy.

As a last thing in this section before we move on to data for rail projects is the cost data for the road projects where such information has been available. As argued in chapter 1, I shall not spend much

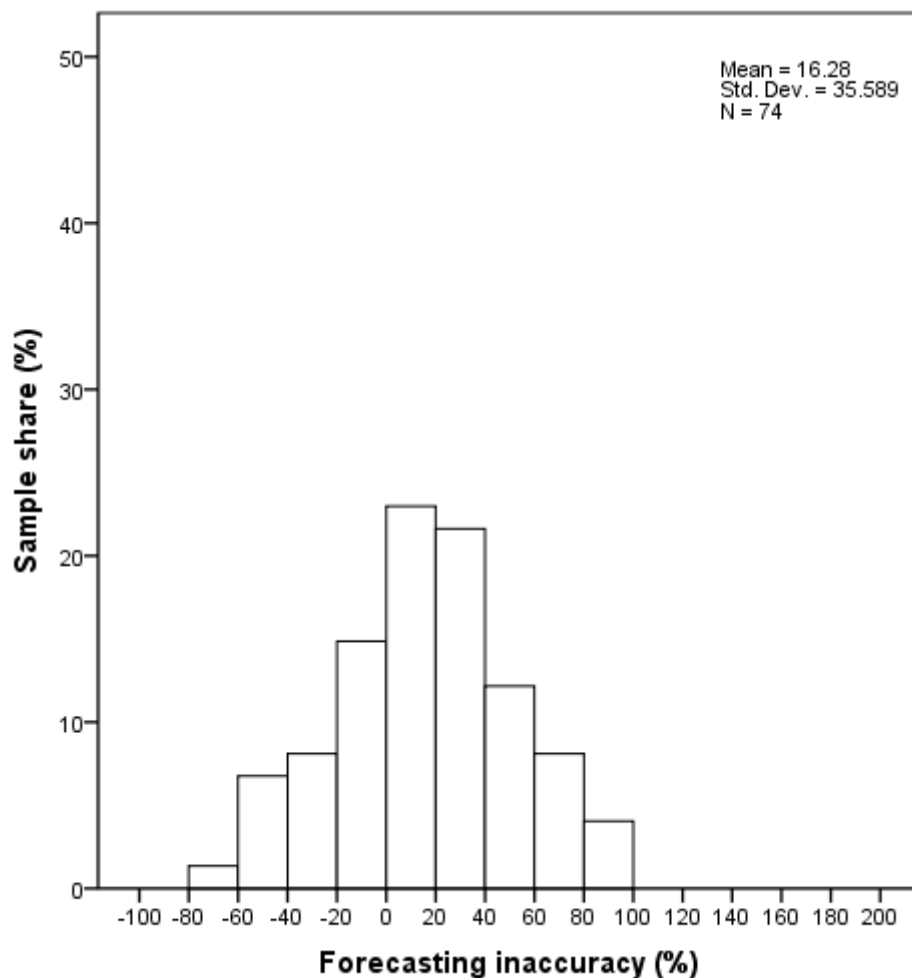


Figure 22: Inaccuracy of cost forecasts for road projects in the total sample.

time discussing the accuracy of cost forecasts, as this is beyond the scope of the present study to make any detailed analysis of such forecasts. I have included data for individual projects only where this has been readily available, and even then it is difficult to assess the validity of the figures. A particular problem lies in knowing which cost items are included in the expected and actual cost figures, as they are rarely presented with much detail in the available decision support documents. I have taken such information into account where available, but often costs for a project will be presented as a single figure in both the associated decision support documents and the available accounting information I have received from the governing directorates. Considering this lack of information, the present data provides the best possible evaluation of forecasting accuracy for costs. The inaccuracy distribution can be seen in Figure 22, where inaccuracy is measured in a similar manner to that of traffic forecasts. All cost figures are of course adjusted for inflation. As was the case for travel demand we witness a slight underestimation of costs, and road projects tend to be roughly 16% more costly than expected. As was the case for traffic forecasts we also observe a considerable spread in forecasting accuracy for costs, indicating a high degree of imprecision.

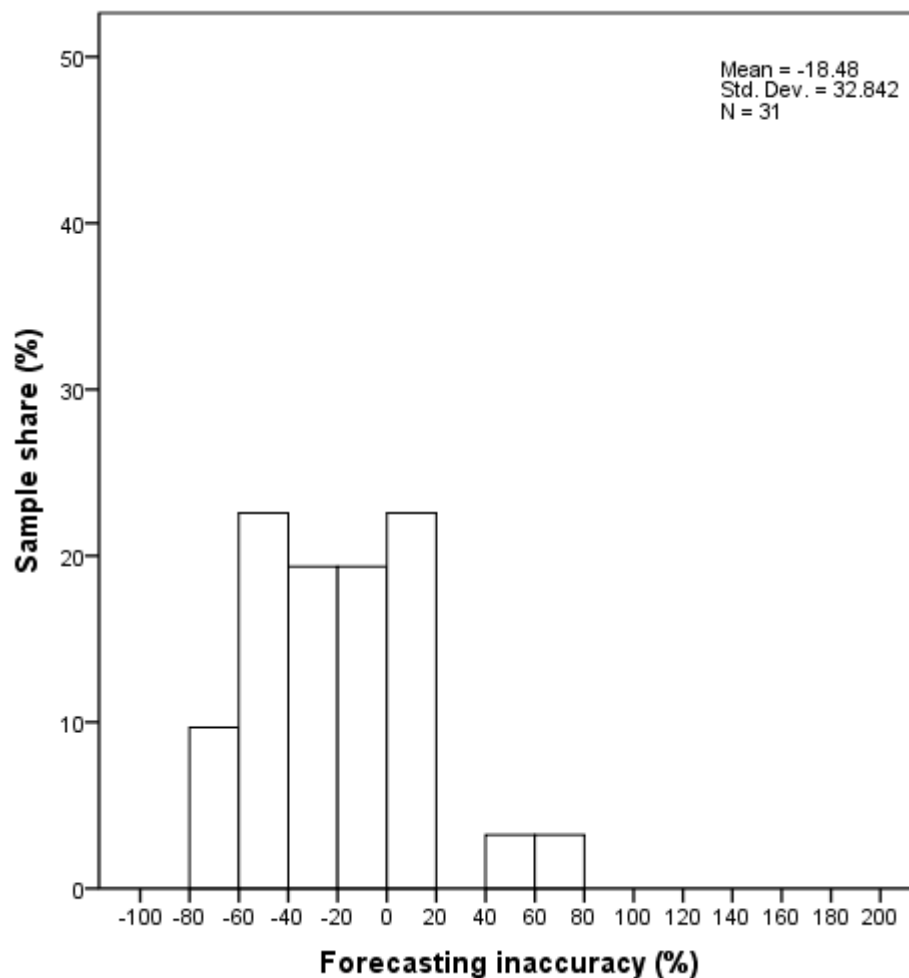


Figure 23: Inaccuracy of traffic forecasts for rail projects in the total sample.

6.1.2 Forecasting accuracy for rail projects

As was the case for road projects, we shall start out by presenting the overall inaccuracy distribution for rail projects to use as a point of departure for our analysis. The previous studies presented in chapter 3 displayed a clear tendency for passenger estimates to be severely overestimated for rail projects, albeit these studies are few³⁴ and their samples relatively small compared with studies of forecasting accuracy for road projects. As was the case for road projects, we notice an approximate normal distribution of forecasting inaccuracy for rail projects in Figure 23. The right sided does not appear as heavy as for road projects, but this could of course merely be a reflection of the smaller sample size. Around two thirds of the projects experience less traffic than forecast, while 42% fall within $\pm 20\%$ of the forecasts. 19% of projects have less than half the forecast traffic volumes, while 32% of projects have between 20-50% less traffic than forecast. The last 7% of projects has over 20% more traffic than forecast. These shares are mainly to give an indication of how the forecasting inaccuracy is distributed, and should not be interpreted beyond that intent. While the sample is

³⁴ While accuracy of demand forecasts for individual rail projects have been the subject of several studies, I am here concerned with studies that have constructed larger samples on similar methodology as the one employed in the present study.

Inaccuracy		Statistic	Std. Error
Mean		-18.48	5.899
95% Confidence Interval for Mean	Lower Bound	-30.52	
	Upper Bound	-6.43	
5% Trimmed Mean		-20.22	
Median		-20.53	
Variance		1078.578	
Std. Deviation		32.842	
Minimum		-66	
Maximum		66	
Range		131	
Interquartile Range		52	
Skewness		.616	.421
Kurtosis		.133	.821

Table 5: Descriptive statistics of forecasting inaccuracy for rail projects.

larger than many other studies in this field, it remains much too small for any sophisticated statistical analysis to be undertaken.

Table 5 displays the most important descriptive statistics of forecasting uncertainty for rail projects to gain a better understanding of the observed distribution in Figure 23. We see that despite the small sample size, the mean is significantly different from zero and shows a clear tendency of forecasts being overestimated. Both the truncated mean and the median are much closer to the full sample mean than was the case for road projects, just as spread, skewness, and kurtosis are also less than for road projects. All of these are due to the more symmetrical shape of the inaccuracy distribution for rail projects, with no extreme outliers in the sample. This is probably because traffic for rail projects is generally an aggregate measure in areas where a substantial commuting population is present. Our inaccuracy measure is therefore less susceptible to extreme fluctuations for rail projects, while some of the road projects are fairly small. Observing extreme values such as double or triple the amount of expected traffic is therefore, *ceteris paribus*, more likely to appear in the road sample than the rail sample.

If we compare the observed distribution with the previous studies of rail projects presented in chapter 3, we find that it is quite different from all of these. Table 6 lists the sample size, mean, and standard deviation for the three previous studies of rail projects to compare with those of the present study. All of the previous studies indicated a drastic overestimation of traffic, with actual passenger volumes around half of the expected figures. The present sample also has a clear tendency of traffic falling short of forecasts, but compared with road projects the bias for rail projects is very different from findings in previous studies. There could be a number of reasons for this, but given the small sample sizes for all studies these can only be considered educated guesses, and I shall only briefly cover two that I consider quite important for this matter. First, the sampled populations do not seem to be as comparable as for road projects, where state motorways dominated the samples

Author(s)	N	Mean	Std. dev.
Pickrell (1990)	9	-65	17
Fouracre et al. (1990)	9	-44	26
Flyvbjerg et al. (2005)	27	-40	52
Nicolaisen (2012)	31	-18	33

Table 6: Comparison of forecasting inaccuracy for rail projects between the present study and the largest previous studies where comparable data exists.

in all studies. For example, around half the projects in Pickrell's (1990) sample are light rail projects constructed during the light rail revival in the US during the 1980s. Prior to these projects was roughly half a century of road orientated development in the US, where no new light rail projects had been built since well before World War II. This must have been a considerable source of uncertainty when preparing forecasts for these projects, since there was a lack of experience at the time regarding commuter responses to such systems. While it is certainly possible that these forecasts were artificially inflated by project proponents, we cannot ignore the special circumstances for these projects. Fouracre et al. (1990) has a sample consisting of metro projects in developing countries, which must also be considered a special type of rail projects. Flyvbjerg et al. (2005) include projects that are more in line with the sample of the present study, but also use both of the aforementioned studies as source material. Consequently, the results are obviously influenced by similar concerns. Second, some time has passed since these results were first published, and the sample in the present study generally consists of newer projects when compared to the previous studies. It is therefore possible that the awareness raised by these previous studies has led to increased scrutiny of forecasts and thereby an improvement of forecasting accuracy. Button et al. (2010) found evidence that demand forecasts had improved in the US after the results presented by Kain (1990) and Pickrell (1990) gained public attention, indicating that awareness of the large uncertainties involved in demand forecasting will itself prompt greater scrutiny of the results. The sample in Button et al. includes Bus Rapid Transit (BRT) systems, and was not included in the review in chapter 3, since this was aimed at rail projects. I have contacted the corresponding author for more information about the projects in the sample, since no details are available in the published article or the associated research report. If the BRT systems do not pose problems for comparison with other rail studies it should probably be included. If they do pose a problem it should be explained that this is the reason for not including this study in a comparison. However, the authors were unable to supply information about the specific projects included.

When comparing the findings in Figure 23 with those of previous studies from chapter 3, we also observe that the distribution in the present study appears to be more centred on the mean than the results presented by Flyvbjerg et al. (2005). This might not be obvious merely through visual comparison with the histograms presented throughout chapter 3, but this is probably just a reflection of my choice to use a uniform scale for these graphs. Smaller histogram brackets would

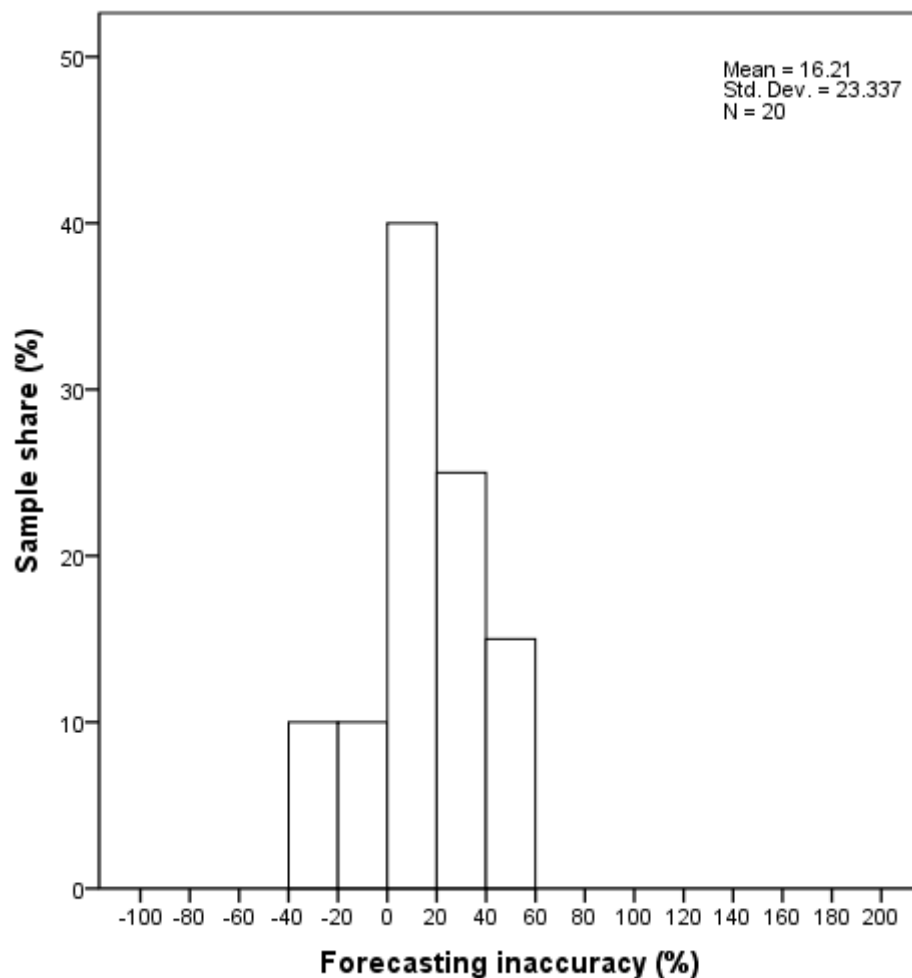


Figure 24: Inaccuracy of cost forecasts for rail projects in the total sample.

reveal a closer fit to a normal distribution, and the information in Table 6 reveals that both bias and imprecision is less for the sample in the present study than those reported in previous studies. This is an interesting finding, since positive outliers inflated the mean in Flyvbjerg et al. (2005), while the present study has an even higher mean in spite of no statistical outliers. Due to the low sample size of Pickrell (1990) and Fouracre et al. (1990) it is difficult to make useful comparisons with the distributions from these samples, but it is clear that bias is severely reduced in the sample from the present study compared to these. However, could these differences be a result of my choice to adjust the inaccuracy measure for mismatching reference points, as was seemingly the case for road projects? Figure 25 presents the distribution of the unadjusted inaccuracy measure for rail projects, and if we compare this with the adjusted inaccuracy from Figure 23 we notice that they are practically identical, both when inspected visually and when considering the mean and spread of the two measures. This is because there has typically been little adjustment required for the rail projects compared to the road projects, and the observed differences between the sample in the present study and those of previous studies can therefore not be explained by the reference point adjustments. Rather, it seems that the populations from which the samples are taken appear to be genuinely different in terms of forecasting accuracy, albeit still displaying considerable inaccuracy in terms of both bias and imprecision.

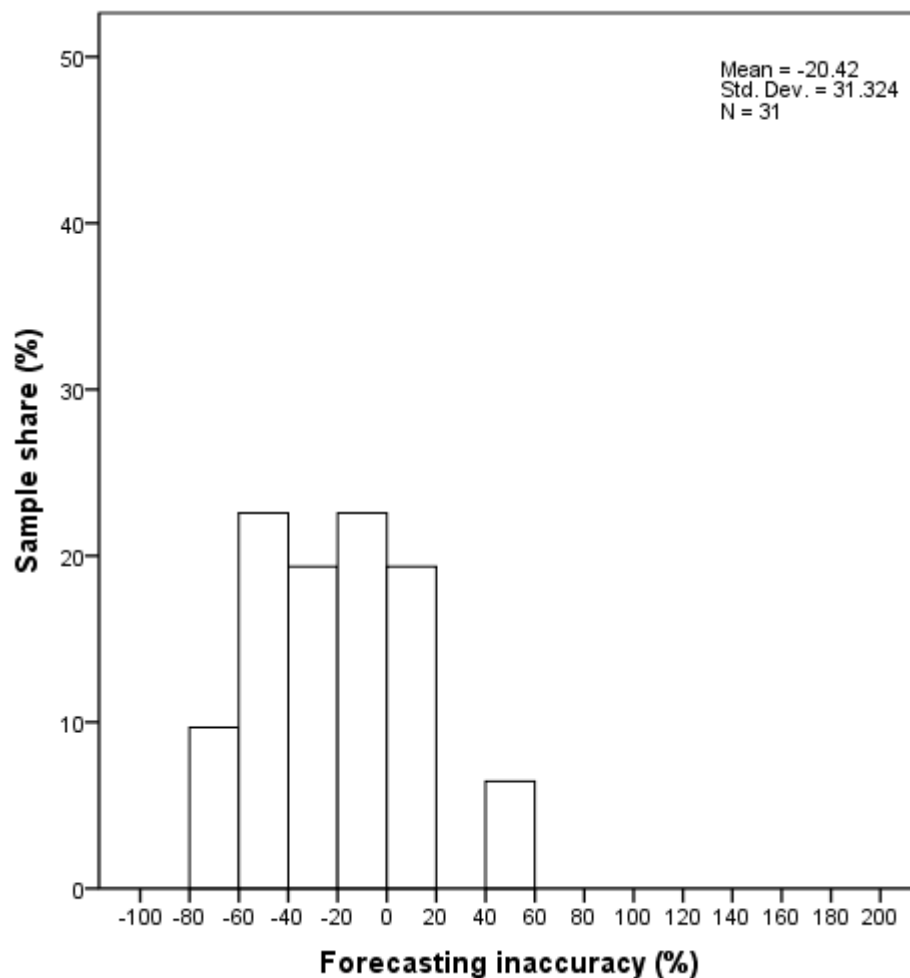


Figure 25: Inaccuracy of traffic forecasts for rail projects in the total sample, unadjusted for mismatching reference points between forecasts and counts.

As with the above section on road projects, I end the section on rail data by briefly discussing the cost data for the rail projects where such information has been available. The limitations regarding data availability are similar for road and rail and I shall not repeat them here, but merely state that it has not been possible to obtain suitable data for all projects³⁵. The inaccuracy distribution can be seen in Figure 24, where inaccuracy is measured in a similar manner to that of traffic forecasts. All cost figures are of course adjusted for inflation. As was the case for road projects we witness a slight underestimation of costs, with rail projects being roughly 16% more costly than expected. As was the case for traffic forecasts we also observe a considerable spread in forecasting accuracy for costs, indicating a high degree of imprecision, although it seems lower than for road projects.

6.1.3 Forecasting accuracy for zero-alternatives

One of the things that have generally been overlooked in previous studies of forecasting accuracy for travel demand is the accuracy of the zero-alternatives with which the proposed projects are compared. As discussed in chapter 5, there are good reasons for this, since it is often difficult to

³⁵ In addition to the results displayed here, cost figures have been obtained for most of the remaining projects in the total sample. Including them does not alter the distribution of inaccuracy at all, but I have chosen to exclude them here as I have not been able to verify the sources of the cost estimates for them.

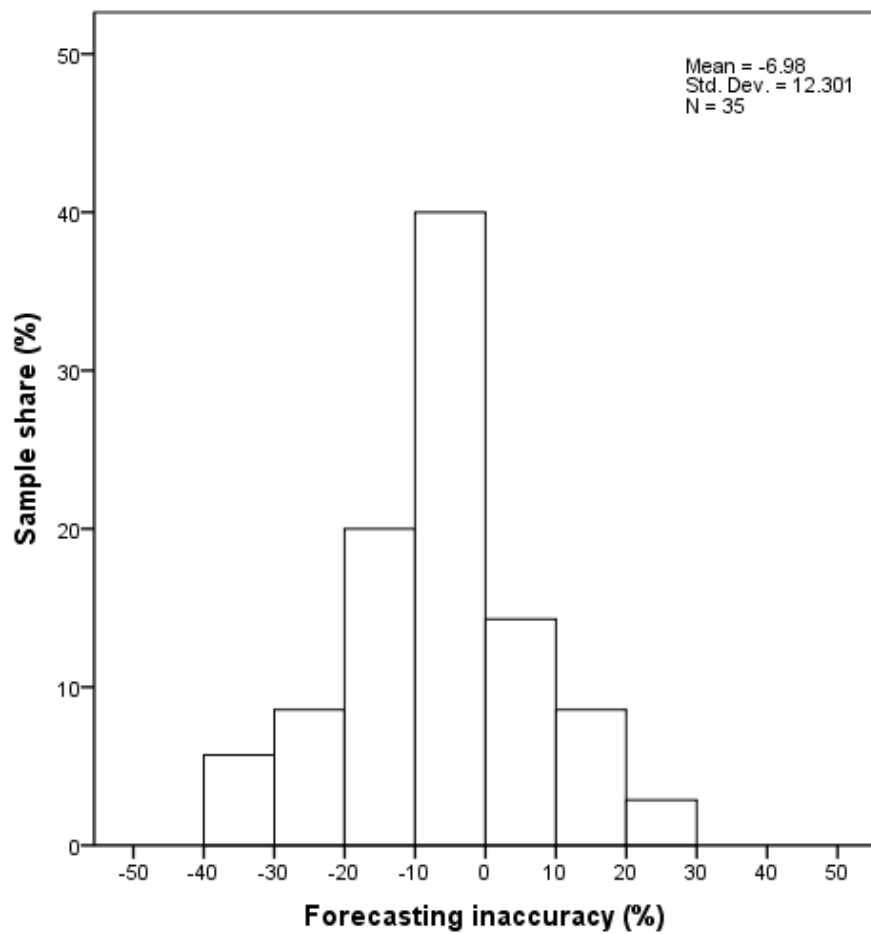


Figure 26: Inaccuracy of demand forecasts for the zero-alternatives for road projects.

obtain data that allows an ex-post evaluation of demand forecasts for the zero-alternatives. In cases where new road projects are built the zero-alternative represents a counterfactual scenario, while the ex-post observations represent a factual scenario. These observations thus reflect the demand levels that result from triggering a set of mechanisms, whereas the forecasts reflect demand levels in situations where these mechanisms are not triggered. To compare the two therefore requires careful considerations of viable proxies for the observed traffic (see section 5.2 for a discussion of how this has been carried out in practice). As a result, the sample of available data for this type of comparisons is much smaller than for the built alternatives presented in sections 6.1.1 and 6.1.2.

Figure 26 displays the forecasting inaccuracy for zero-alternatives for the road projects where such data has been available. None of the rail projects in the sample had data available that allowed a similar comparison, and I shall therefore focus exclusively on road projects in the treatment of forecasting accuracy for zero-alternatives. In contrast with the built alternatives we observe a slight overestimation of traffic for the zero-alternatives. Table 7 displays the most important descriptive statistics of forecasting uncertainty for zero-alternatives, where we observe that the mean is significantly different from zero despite the small sample size. Of interest is also the median value, which indicates that a sizable majority of projects are overestimated. This is due to a very small spread of the distribution compared to the measures for built alternatives, as well as a noticeable

		Statistic	Std. Error
Inaccuracy	Mean	-6.98	2.079
	95% Confidence Interval for Lower Bound	-11.21	
	Mean Upper Bound	-2.76	
	5% Trimmed Mean	-6.85	
	Median	-5.84	
	Variance	151.311	
	Std. Deviation	12.301	
	Minimum	-34	
	Maximum	21	
	Range	56	
	Interquartile Range	16	
	Skewness	-.182	.398
	Kurtosis	.252	.778

Table 7: Descriptive statistics of forecasting inaccuracy for zero-alternatives for road projects.

lack of outliers in the sample. Intuitively this seems quite logical, since the zero-alternatives represent a situation where no changes are introduced to the existing system. As a result, the many uncertainties related to behavioural changes in response to system changes do not apply here, and in many cases extrapolation of current trends will probably serve as a fairly accurate prediction of future development.

Such simple forecasting methodology appears to be standard practice for constructing the zero-alternative in many decision support documents, where an annual growth rate is typically used to extrapolate traffic volumes towards a future reference year based on average growth trends over the last decade or so. However, such an approach does not take into account the detrimental effects on demand that follow from increased levels of congestion. If traffic does indeed grow at a steady rate, whether it is linear or exponential, system capacity will reach its limit at some point, and additional growth in traffic will start slowing down as potential users either seek other modes of transport or do not travel at all. This is likely to be the reason why we observe a systematic overestimation of traffic levels in the zero-alternatives, where more than three fourths of road projects experience less traffic than expected in a situation where the proposed project is not built.

As was the case with forecasting accuracy for built alternatives, there is a small but systematic bias in the appraisal of zero-alternatives. However, while these biases might not appear to be very severe when compared to the large imprecision associated with forecasting in general, there are at least two good reasons why we should pay attention to them. First, both biases distort the overall appraisal of road projects in the same direction, and need to be considered in combination since benefits are calculated by comparison between the two types of forecasts. The aggregate forecasting inaccuracy thus results in around 18% additional traffic for road projects when compared to predicted figures. Second, both measures are related to a single target year, which is usually sometime in a foreseeable future relative to when the forecast was made. If demand is

overestimated in the short-term for the zero-alternatives as a result of ignoring the demand reducing effects of congestion, chances are that this bias will only increase exponentially for forecasts that cover a longer timeframe. Taken together, these two arguments indicate that appraisal of road projects severely underestimate the adverse environmental impacts of additional road capacity in the long term. For the same reasons, initiatives to reduce road capacity might be systematically underestimated in their ability to reduce adverse environmental impacts.

6.2 Exploring common project traits

In the previous section we observed the inaccuracy distributions for forecasting accuracy of both road and rail projects. In line with previous research findings we found demand for road projects to be slightly underestimated on average and demand for rail projects to be more considerably overestimated on average. There thus appear to be significant bias in demand forecasting for both road and rail projects, although the bias for rail projects seems less than what has been found in other studies. However, while the distinction between road and rail projects has proved a rough estimator for the bias of the mean, forecasts were also found to be fairly imprecise in general, indicated by large standard deviations for both project types. Additional particulars of individual transport infrastructure projects must therefore be taken into account when assessing the forecasting accuracy of demand, as there are clearly large deviations from the observed means. In the following we shall explore some of the possible categories (reference classes if you will), that might help in explaining the observed inaccuracies. While this will be done in a semi-exploratory way, the categories we wish to consider are of course quite theory-informed, and I shall briefly present why I have considered it of interest to consider the categories presented herein.

6.2.1 Forecasting accuracy over time

Some studies have suggested that we ought to expect an improvement in forecasting accuracy for newer projects, since more sophisticated forecasting techniques have been made available over time and knowledge in this area has improved (e.g. Flyvbjerg 2007; Pickrell 1992). However, this has not always been found to be the case, which has led some authors to conclude that improved models are not of particular concern if the goal is more accurate forecasts. While I agree that it is important to conduct time series analysis, I am more cautious about dismissing the role of model improvements merely by the use of such data. Such a conclusion implicitly assumes that applied modelling techniques have become more sophisticated over time, and not just that improved techniques have been developed. This might be plausible, but it does not appear to have been investigated very thoroughly in the existing body of literature. In addition, the conclusion also assumes that the older projects are as complex as newer projects, which is certainly much less plausible. Transport infrastructure networks are becoming increasingly complex in larger urban areas due to many new and competing services, widespread introduction of information technology, and other changes to commuting behaviour. I shall return to this topic later in the present chapter, but for now it seems sufficient to advise against any temptation towards over-analysing the data presented here, as it can merely tell us if forecasting accuracy has improved or not. Any discussion of possible causal mechanisms must require additional data or convincing external references to be of much use. From the sparse information on modelling that has been available in the decision support documents obtained for the present study, there is no support for large advances in the sophistication of models used in practice.

Figure 27 shows a scatterplot of forecasting inaccuracy and opening year for road projects. We observe that there appears to be little to no correlation between the two, indicating that forecasts have not become more accurate over time. If anything, we notice a slight increase in the mean inaccuracy over time, indicating that underestimation of traffic is more common among newer projects. It should be noted that the majority of the projects have been completed in the new millennium, but even when considering this sample bias there appears to be no indication of forecasts accuracy changing over time. Flyvbjerg et al. (2005) observed that Danish road projects in their sample seemed to suffer from assumption drag, but there appears to be no clear evidence in support of this in the present study (see Figure 28). Norway is the only country where inaccuracy appears to differ over time, but this is merely due to inflation from two extreme cases that were opened around the same time. Using the year that the forecast was produced as a reference point rather than the opening year of the project does not reveal any improvements in forecasting accuracy over time (Figure 29). If we sort the projects by opening in intervals of 5 years we notice that general forecasting accuracy does appear to increase quite dramatically for newer road projects, since a box plot reveals a much closer concentration on zero for the period 2005-2009 than previous intervals (Figure 30). However, we also observe a great deal of outliers in this period, which indicates that bias is still a considerable problem among certain projects.

Figure 31 shows a scatterplot of forecasting inaccuracy and opening year for rail projects. Here we observe a noticeable reduction of bias over time when compared with the data for road projects, although imprecision remains high. While the sample is small, opening year alone is able to explain around 20% of the observed forecasting inaccuracy, even if considering the single project opened prior to 1990 as an outlier (although the raw coefficient obviously drops by doing this). If we use forecast year rather than opening year the tendency appears even stronger (Figure 32). This is in line with findings by Button et al. (2010) for US transit projects, although we cannot undertake the same analysis of a potential 'Pickrell effect' as done in that study due to lack of data for projects completed prior to 1990. If we compare the findings of the present study with earlier projects from Flyvbjerg et al. (2005) it seems that a similar argument could possibly also be made for Scandinavian and UK rail projects, but this is somewhat speculative since we do not know whether their sample is actually representative for such a purpose. The authors include projects from all four countries used in the present study, but do not provide any information that allows us to gauge the representativeness of the overall results for these specific countries. No statistically significant differences were found between geographical areas by Flyvbjerg et al. (2005), but since the sample covers only 27 projects spread on 14 different countries this would also be surprising, and I agree with the authors' advocacy for more data to be collected for in-depth studies of this. I have chosen to exclude figures for the country-wise breakdown since it does not make much sense due to lack of sufficient projects in each category.

From the available data in the present study there does not appear to be any clear improvement in forecasting accuracy over time for road projects, although there is some indication that demand for many projects were consistently underestimated for projects opening during the 1990s. This bias is less evident for projects completed after 2000, although there are still projects opening in this period that attract considerably more traffic than expected. For rail projects, the data is more fragmented due to the smaller sample, but there is some indication that underestimation of demand is less of a

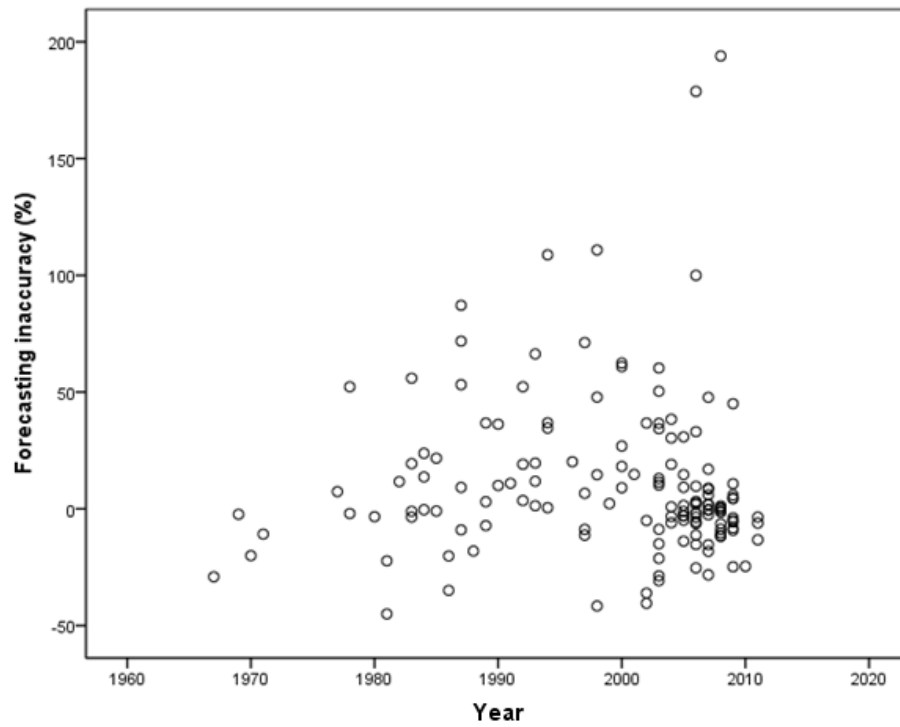


Figure 27: Scatter plot of forecasting inaccuracy and opening year for road projects.

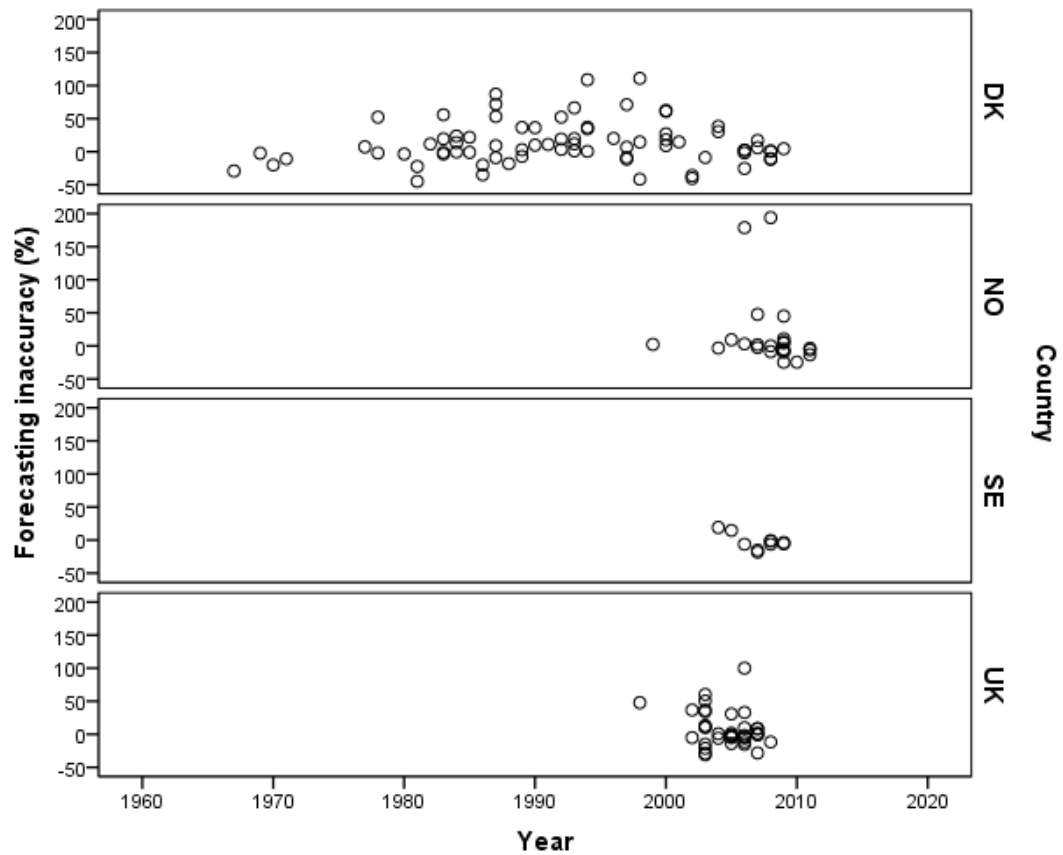


Figure 28: Scatter plot of forecasting inaccuracy and opening year for road projects, split on individual countries.

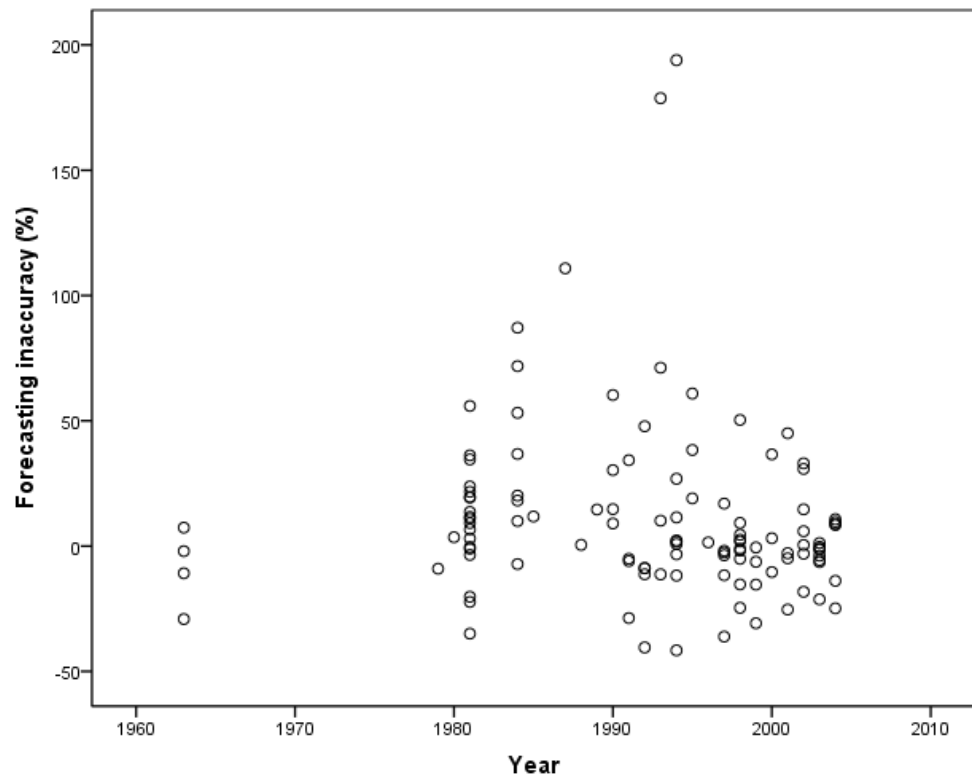


Figure 29: Scatter plot of forecasting inaccuracy and forecast year for road projects.

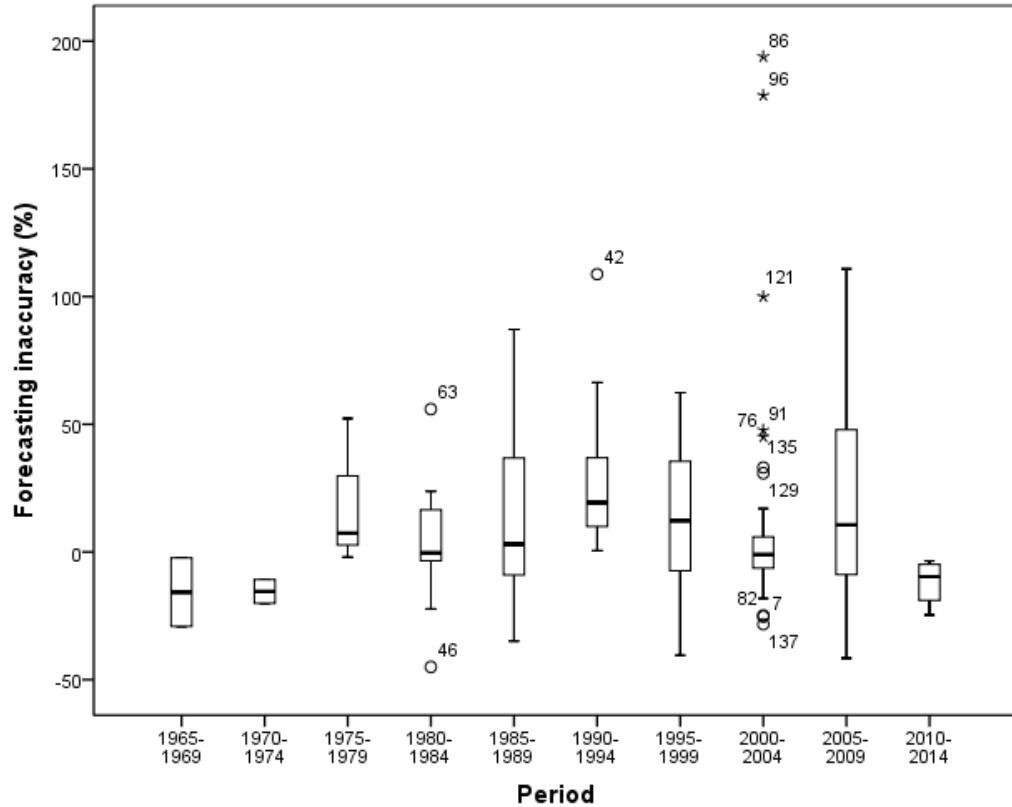


Figure 30: Boxplot of forecasting inaccuracy and opening year for road projects, grouped in 5 year intervals.

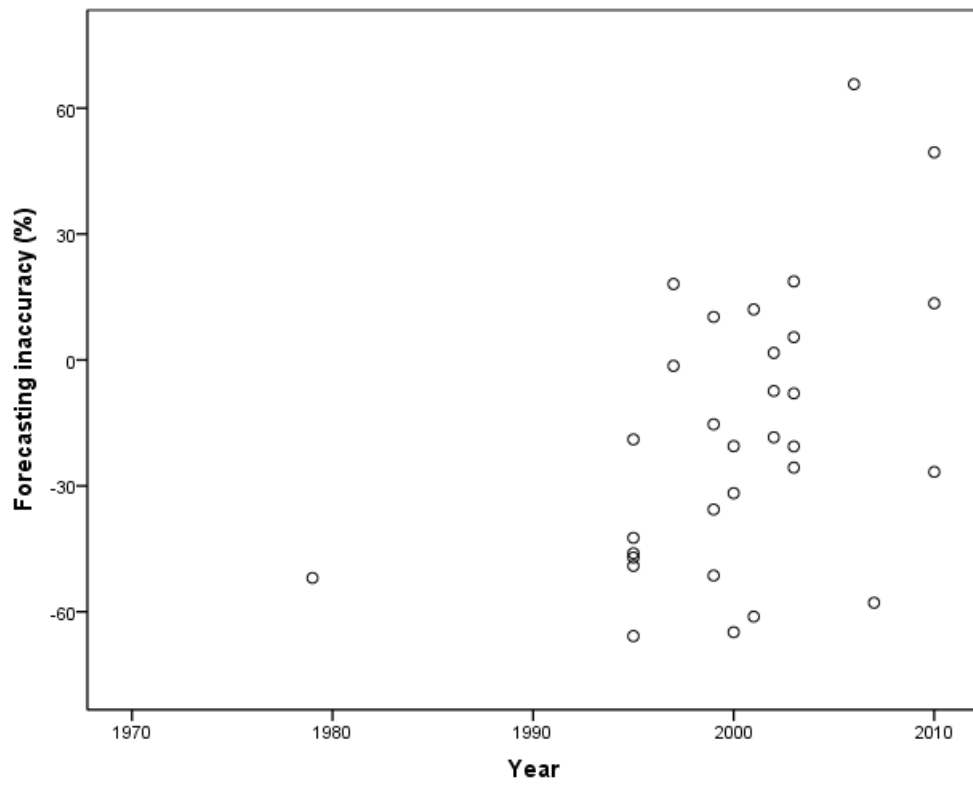


Figure 31: Scatter plot of forecasting inaccuracy and opening year for rail projects.

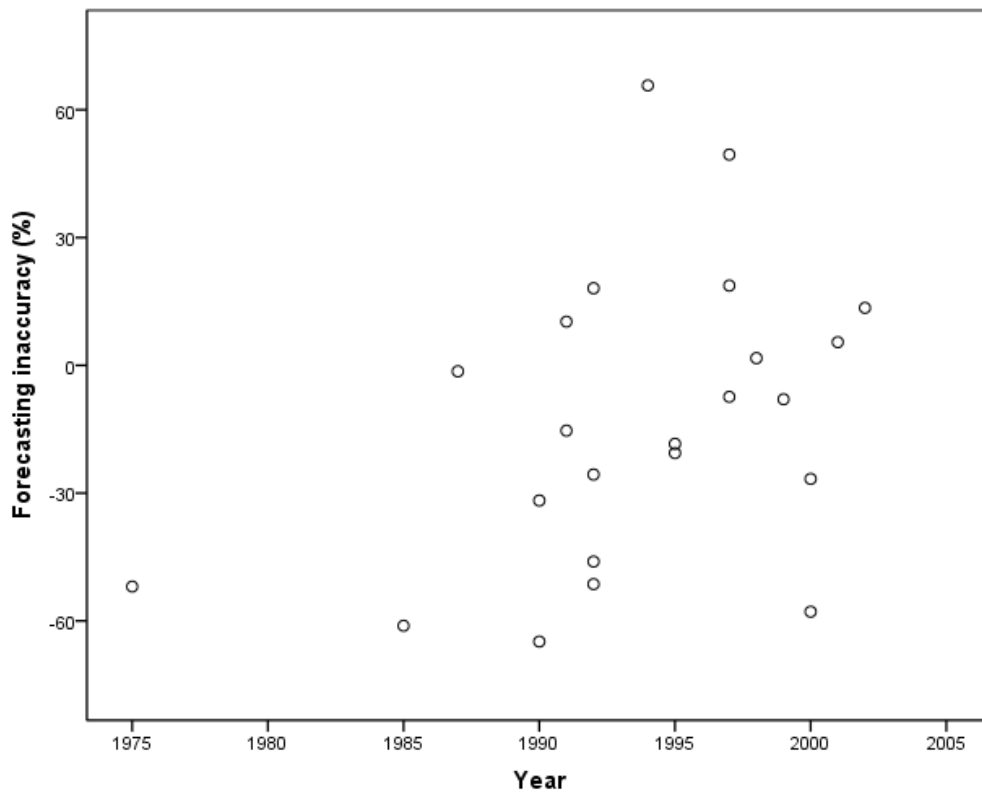


Figure 32: Scatter plot of forecasting inaccuracy and forecast year for rail projects.

problem for newer rail projects than for older. However, both road and rail projects display a considerable degree of imprecision, which makes it difficult to point to any consistent trends in forecasting accuracy over time. The most extreme deviations from expected figures are found among relatively new projects for both road and rail projects.

6.2.2 Short vs. long term forecasting

Another factor we would expect to be an indicator for forecasting accuracy is the time span we are forecasting ahead, since the temporal uncertainty is obviously greater for long term forecasts than for short term forecasts. Næss and Strand (2012) have argued that traditional transport models are inherently poorly equipped for dealing with forecasts covering a particular long time span, and while they do not refer to the taxonomy for uncertainty proposed by Walker et al. (2003), there does appear to be many similarities in the argumentation surrounding non-quantifiable uncertainties. Figure 33 displays a boxplot of forecasting inaccuracy for road projects and the span between forecast year and opening year. There does not appear to be any strong evidence that forecasts far into the future are considerably more inaccurate than short term forecasts in general, but it is clear that the largest deviations from expected figures are found among long term forecasts. A scatterplot³⁶ indicates a similar tendency for rail projects (Figure 34), where we also observe that the largest deviations are found among forecasts covering the largest time span.

The distinction between what counts as short and long term forecasts is of course fairly ambiguous here, and most of the projects in the sample of the present study would probably be associated with long term forecasting, if we take this to mean anything looking more than two or three years ahead. This distinction is merely a matter of taxonomy, and what should be clear from the data is that there is considerable inaccuracy to be found among all forecasts prepared for reference scenarios that lie more than two years ahead, and that there is a risk of quite dramatic inaccuracies when reference scenarios lie more than eight years ahead.

6.2.3 Project types

So far we have treated road and rail projects as two separate but homogeneous groups, but among each of them are of course quite different types of individual projects. We would not intuitively expect these to be associated with the same types of uncertainties, or at least not to the same degree. A widening of an existing motorway is inherently a less complicated forecasting challenge than a new fixed link that replaces a ferry. The former allows additional traffic on an existing network, where demand is typically already reaching maximum capacity (hence the reason to add lanes). Conversely, the latter drastically reduces travel time between two regions that previously enjoyed only limited connectivity, and might result in commuting patterns that are very different from the situation without a fixed link. I have crudely split the road projects into four different categories of project types (MWU, MWC, BPC, and FXL). MWU covers motorway or highway projects that mainly consist of upgrades to an existing link, such as a widening or junction improvement. MWC covers new motorway or highway projects that are constructed between regions that previously had no direct connection or where the new link is a considerable upgrade to existing links

³⁶ A boxplot has not been included for rail projects since the sample is too small for any useful information to be conveyed in such a figure.

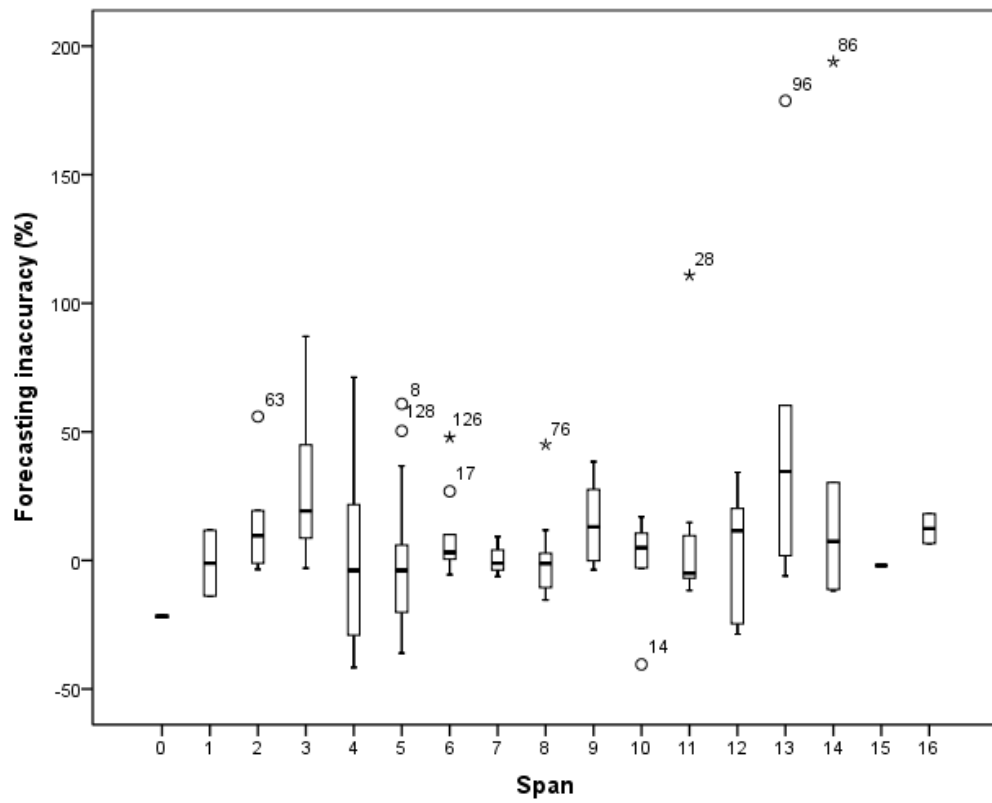


Figure 33: Boxplot of forecasting inaccuracy for road projects and the span between forecast year and opening year.

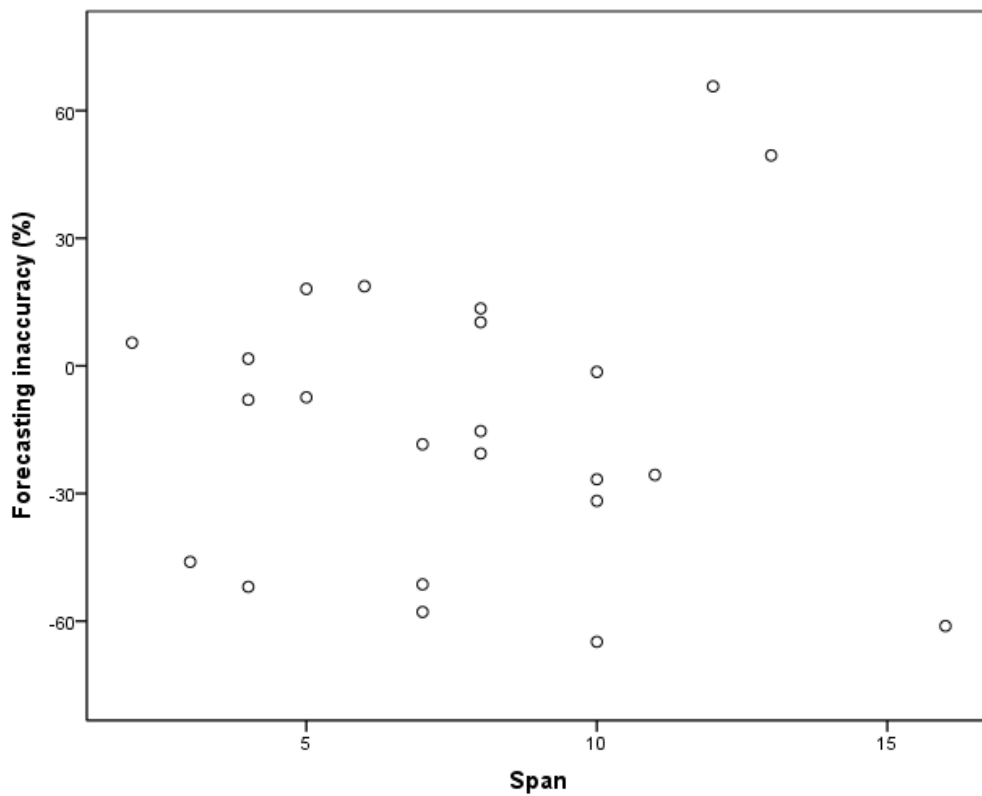


Figure 34: Scatter plot of forecasting inaccuracy for rail projects and the span between forecast year and opening year.

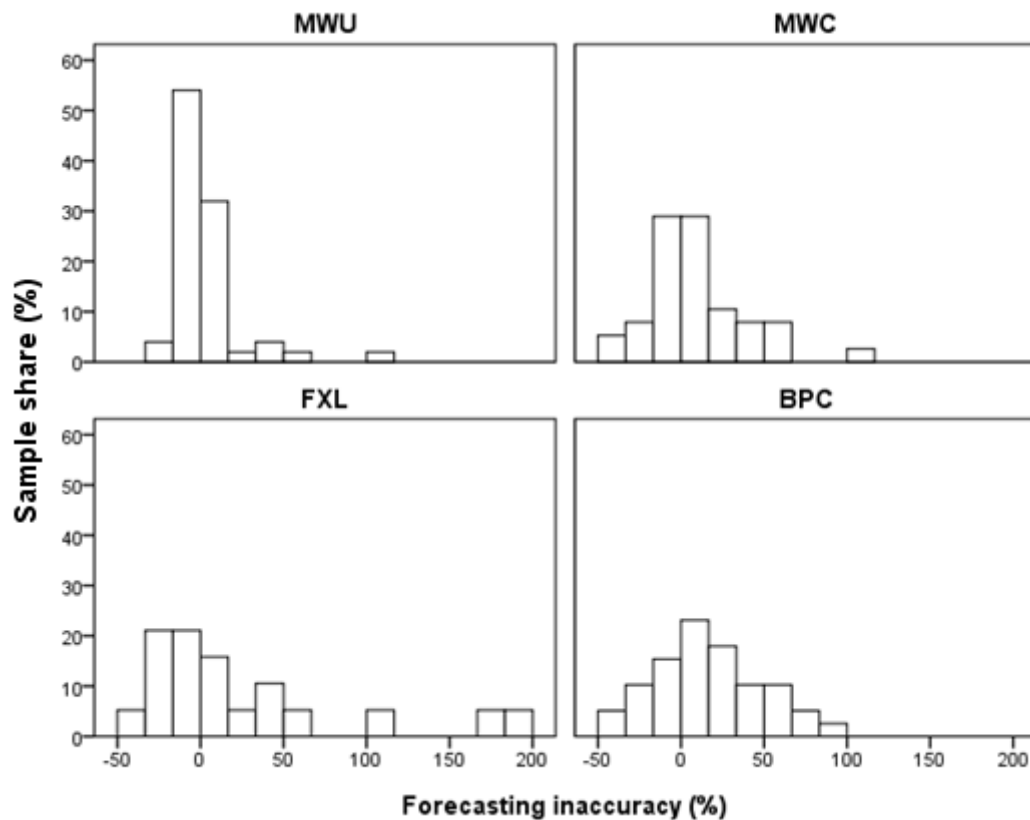


Figure 35: Inaccuracy of demand forecasts for road projects, split on different project types. MWU = motorway/highway upgrade, MWC = motorway/highway construction, FXL = fixed link, BPC = bypass construction.

Project type	Mean	Std. Deviation	N
Bypasses	17	32	39
Fixed links	28	66	19
Motorway constructions	9	29	38
Motorway upgrades	2	20	50
Total	11	35	146

Table 8: Inaccuracy of demand forecasts for road projects, split on different project types. MWU = motorway/highway upgrade, MWC = motorway/highway construction, FXL = fixed link, BPC = bypass construction.

of a lower road classification. BPC covers bypass projects that aim at redirecting traffic around urban areas. FXL are new fixed links (tunnels/bridges) that connect areas that previously required a ferry or a substantial detour to reach.

Figure 35 displays the inaccuracy among the four different types of road projects. Here we observe that upgrades to existing links (MWU) typically experience fewer problems with forecasting inaccuracy than other project types, with traffic on the vast majority of these projects deviating less than $\pm 20\%$ from expected values. New motorway or highway projects (MWC) also has most of the projects falling within this span, but here we notice a greater tendency of deviation, which is biased towards a general underestimation of traffic. Forecasts for fixed links (FXL) and bypass constructions

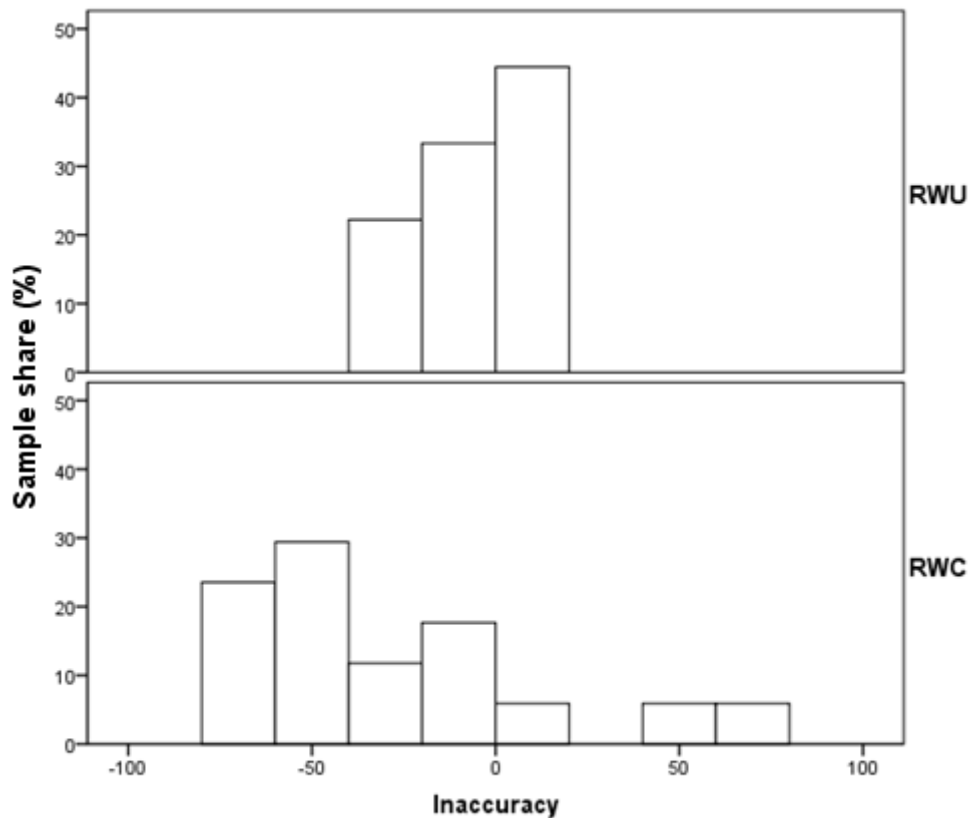


Figure 36: Inaccuracy of demand forecasts for rail projects. RWU = railway upgrade, RWC = railway construction.

Project type	Mean	Std. Deviation	N
Railway constructions	-27	40	17
Railway upgrades	-5	17	9
Total	-18	33	26

Table 9: Inaccuracy of demand forecasts for rail projects, split on different project types. RWU = railway upgrade, RWC = railway construction.

(BPC) appear much more uncertain, with no clear concentration to be observed in the distributions of inaccuracy for these project types. From Table 8 we gain further confirmation that both bias and imprecision is negligible for MWU projects, slightly more concerning for MWC projects, and quite dramatic for BPC and FXL projects.

Figure 36 and Table 9 display similar data for rail projects. Due to the small sample size for rail projects I have only split these into two very crude categories (RWU and RWC), since the individual categories would otherwise end up with too few projects for analysis. RWU covers upgrades to existing links, such as additional tracks or new stations. RWC covers the construction of new connections or substantial upgrades that include new services (e.g. the introduction of high speed rail). As was the case for road projects, we observe that upgrades to existing links typically experience much fewer problems with forecasting inaccuracy than construction of new links or services. Table 9 confirms that both bias and imprecision are much greater for RWC than for RWU projects.

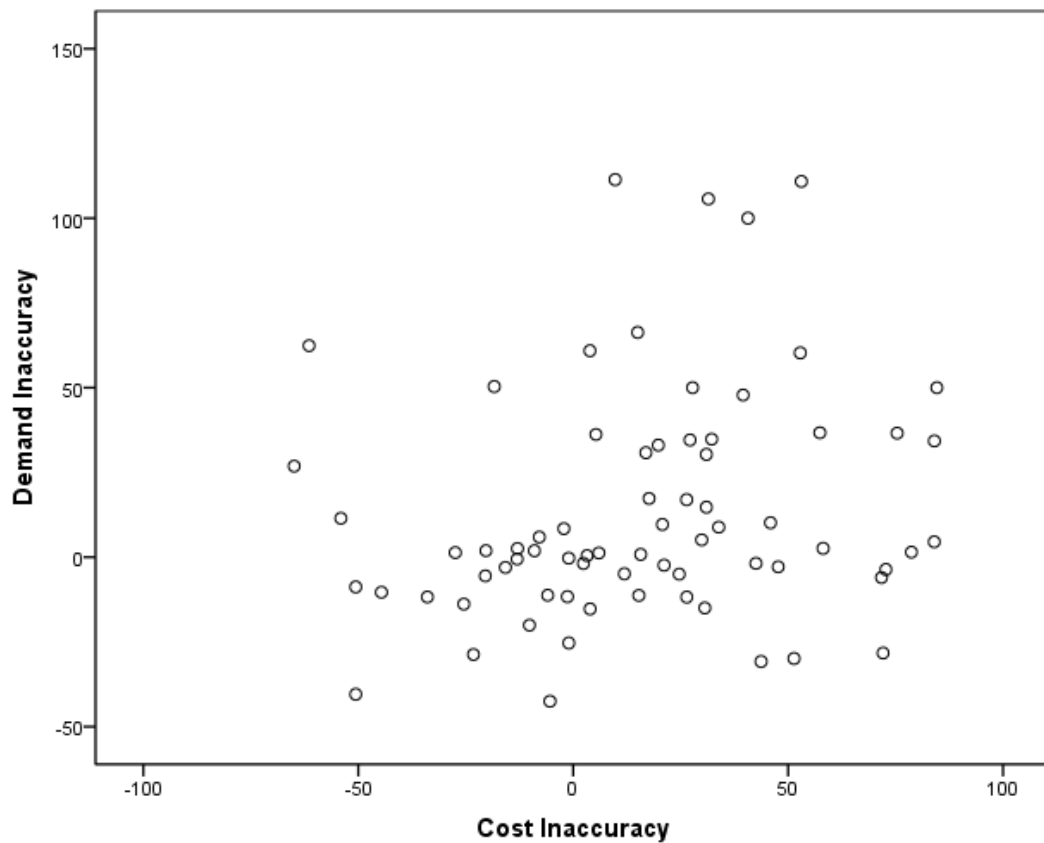


Figure 37: Scatterplot of forecasting inaccuracy of demand and cost for road projects.

6.2.4 Cost inaccuracy

Some authors have speculated that moral hazard among project proponents can be a major source of forecasting inaccuracy (e.g. Flyvbjerg, Holm, and Buhl 2005; L. R. Jones and Euske 1991; Kain 1990; Næss 2011; Wachs 1989). This argument is motivated by claims that the observed forecasting bias for both demand and costs has a tendency to favour project implementation by either overestimating benefits or underestimating costs. Perhaps the most direct accusation of this sort comes from Flyvbjerg's (2007) hypothesis that there is a 'survival of the unfittest' among projects, and that those with the most polished figures are ultimately the ones that get approved. This is obviously a quite dramatic interpretation of biased forecasts since there could be other reasons for bias to occur, but it would likewise be foolish to disregard the explanatory power of strategic behaviour. Furthermore, this is not an easy hypothesis to test, as is also evident from the lack of convincing evidence in previous studies.

However, if strategic behaviour is really a major source of inaccuracy, we would expect forecasts for demand and costs to be somewhat correlated. While cost data has not been a prime concern in the present study, the available data at least allows us to investigate this for part of the sample. Figure 37 displays a scatterplot of the accuracy of demand and cost forecasts for road projects. We observe no particularly convincing relationship between these two variables, but statistical tests show a small positive correlation that is statistically significant at the 10% level. However, the correlation is much too small to support a hypothesis of strategic misrepresentation as a main explanatory power of

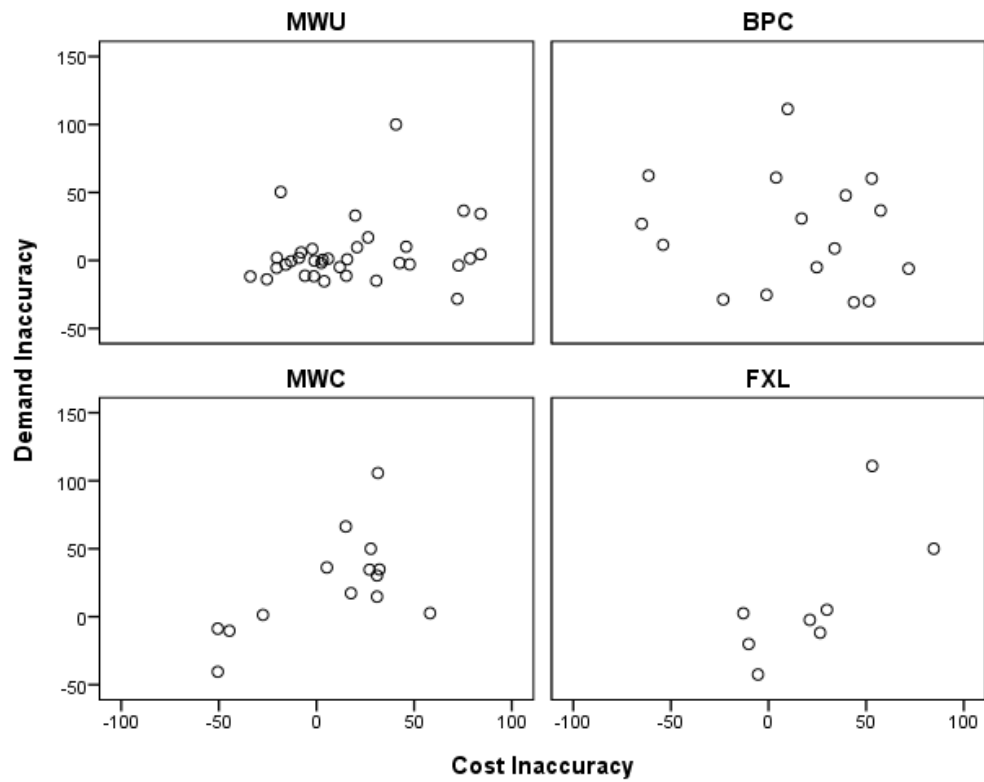


Figure 38: Scatterplot of forecasting inaccuracy of demand and cost for road projects, split on individual project types. MWU = motorway/highway upgrade, MWC = motorway/highway construction, FXL = fixed link, BPC = bypass construction.

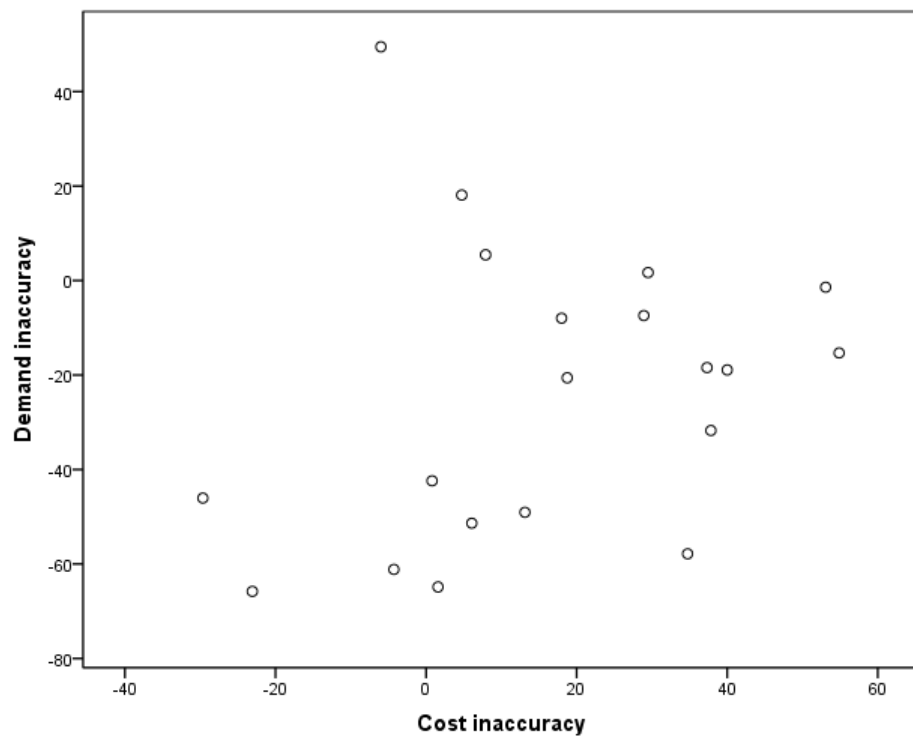


Figure 39: Scatterplot of forecasting inaccuracy of demand and cost for rail projects.

forecasting inaccuracy, unless proponents deliberately choose to only misrepresent one type of forecasts. It should be noted that the incentive for misrepresenting demand for road traffic is not always unidirectional, which could explain why statistical tests show very weak correlation. However, Figure 37 does not indicate that the numerical measure of inaccuracy is particularly correlated with cost inaccuracy either. Since the incentive for misrepresenting costs must be assumed to be unidirectional this indicates that if any strategic misrepresentation of demand and costs are taking place, it is certainly not coordinated between the people responsible for these respective forecasts.

It is of course possible that correlation patterns might be identified for certain subgroups of projects, and Figure 38 displays a scatterplot for the individual project types introduced in section 6.2.3. While lack of verified cost data leaves some of the categories with relatively few items, there are statistically valid correlations at the 5% level to be found for MWC and FXL projects. On top of this the coefficients of correlation are 5-10 times greater than for the whole sample, just as the coefficients of determination are 10-15 times greater. This suggests that demand and cost inaccuracies are much more correlated for new construction of motorways and fixed links than for other types of road projects, and perhaps more in line with the hypothesis of strategic misrepresentation. However, would we also expect them to be positively correlated for these project types? Fixed links are often funded via user charges, and thus an incentive exists for presenting sufficient demand for the project to be feasible. In case of projects that are free of charge we would still expect high demand to be an influential factor in decision-making, since the user benefits primarily comes from previous ferry users and induced traffic. In this context we would therefore expect a negative correlation between demand and cost inaccuracies if strategic misrepresentation is a major influence, rather than a strong positive correlation that we are instead observing for FXL projects in Figure 38. Incentives for misrepresenting demand on new motorways is probably a lot less unidirectional, since additional demand could increase or decrease estimated user benefits depending on the level of congestion in the network. However, for many of the environmental effects there is a clear incentive from project proponents to underestimate demand, so the adverse effects appear to be less than what is actually going to be the case. This goes well in line with what we observe for both MWC and FXL projects in Figure 38.

The correlations we observe for individual project types thus appear to give some support of a strategic misrepresentation hypothesis for certain project types. However, the samples are still relatively small for these subgroups, and there could be perfectly reasonable explanations for this correlation that do not include any deceptive behaviour. For example, if a project is built with three lanes in each direction rather than two due to design changes after a decision to build has been taken, the project becomes both more expensive but also able to carry more traffic. Conversely, if we reduce the size of a project we also reduce the costs and the ability to carry traffic. In both cases we would observe a positive correlation, as is the case in Figure 38. Should we consider such examples as cases of strategic misrepresentation? After all, policy makers are typically informed of the impacts on costs and demand from such design changes prior to agreeing with them, and have thus acted on the basis of this new assessment rather than just the original forecast. However, if there is a general tendency for projects to undergo considerable design changes after a decision to build has been taken, this could be seen as an attempt from project proponents to get the projects through an initial approval phase, since they know that sunk costs prevent necessary design changes in being adopted

at a later stage. This would in itself be an interesting analysis, but such inquiry lies beyond the scope of the present study. However, the thought experiment hopefully illustrates the difficulties involved in assigning a sinister motive to stakeholder actions based solely on a statistical comparison of expected and actual outcomes.

Figure 39 displays a similar cost comparison for rail projects as was presented for road projects in Figure 37. The results are similar to those for roads, and there is no significant correlation to be found. The small sample size makes it pointless to try and split the analysis on subgroups of different project types, and I shall not devote more attention to the cost estimates here.

6.2.5 Qualitative assessment

In this section I shall try to identify possible characteristics of projects with different levels of accuracy in a more explorative manner. So far we have seen no particular strong determinants among the quantitative assessments, but analysis at the individual project level might reveal information that is lost in the aggregate figures. I shall not describe all projects in detail in the following, but only highlight a few illustrative examples of projects in the different spans of accuracy. The assessment is not particularly structured, since it is largely based on notes and observations I have made while gathering data for individual projects. This does not mean that no information of importance can be obtained from such analysis, and I personally consider such fragmented data to be highly valuable for exploratory investigations. Given the lack of large-scale studies of forecasting accuracy it might even be considered necessary, since there are few comparative studies in which to anchor such analysis. A summary of the observations can be seen in Table 10, while more detailed examples will be provided in the remainder of this section.

We start by looking at the most extreme cases, where ex-post traffic volumes are above 200% of forecasts, of which there is a total of four road projects. Three of these are fixed links, of which two are smaller tunnel projects in Norway (Eiksundtunnelen and Fedatunnelen) and the last the Great Belt Bridge in Denmark. The demand forecasts are generally not treated with any particular interest in the available decision support for the two Norwegian projects, and for one of them I had to resort to the national transport plan rather than the decision support to obtain any actual figures on expected traffic volumes. Both projects seem to be motivated by a desire to improve links between areas that were previously poorly connected, since reporting of economic appraisals of the expected demand is not addressed in much detail for either project³⁷. If this lack of reporting on economic effects is an indication of such effects being largely irrelevant to a decision on whether to build or not, the large inaccuracy in the associated forecasts could possibly be explained by lack of resources devoted to more thorough analyses.

In the case of the Great Belt Bridge it has been quite difficult to point to a specific reference point for which forecast to use, since this project has been appraised several times throughout the 20th century. Initial drafts for a fixed link go back more than 150 years, while more serious plans began to

³⁷ This is not to say that project feasibility has not entered the public debate in these cases. For example, Eiksundtunnelen has been used as an example of heavy subsidies to outskirt areas in Norway due to the low travel demand in the area (DN 2009). However, the observed forecasting inaccuracy can only have improved the feasibility of the project in this case, and the debate seems to be a more principal matter of whether such indirect subsidies are at all justifiable.

Forecasting accuracy		Description of typical project types
ROAD	Highly underestimated ($I > 50$)	<ul style="list-style-type: none"> Fixed links that attract and generate a drastic amount of travel, typically when replacing ferry lines. Motorways and bypasses that have been assessed based on national trends, but where local trends differed substantially.
	Underestimated ($50 > I > 20$)	<ul style="list-style-type: none"> Motorways and bypasses that have been assessed with little or no concern for induced demand. Motorways and bypasses that have been assessed with relatively small areas of influence.
	Fairly accurate ($20 > I > -20$)	<ul style="list-style-type: none"> Upgrades to existing links where demand is very similar before and after project implementation. Motorways in peripheral areas with relatively modest amounts of existing traffic and no major urban hubs.
	Overestimated ($I < -20$)	<ul style="list-style-type: none"> Motorways and bypasses that have assumed most traffic in a corridor to switch to the new link. Forecasts where demand depends on additional stages or separate schemes that end up being delayed or scrapped.
RAIL	Underestimated ($I > 20$)	<ul style="list-style-type: none"> New rail projects in major urban areas with good connections to additional transit services.
	Fairly accurate ($20 > I > -20$)	<ul style="list-style-type: none"> Upgrades to existing links where demand is very similar before and after project implementation.
	Overestimated ($-20 > I > -50$)	<ul style="list-style-type: none"> Inter-city connections for long commutes.
	Highly overestimated ($I < -50$)	<ul style="list-style-type: none"> Inter-city connections that expect to capture large shares of the total commuting in the corridor.

Table 10: Overview of the most typical project types observed in various spans of forecasting accuracy.

shape when the Danish ministry of public works (now the ministry of transport) started assessing possible solutions in 1948. The first law was passed in 1973, but two oil crises during the 1970s made many people sceptical of the future of private cars and it was not until 1987 that the final construction law reached agreement in parliament. An abundance of reports has been published both prior to and after the construction law was passed, and they differ quite substantially in their predictions of travel demand. I have thus chosen to base my assessment of forecasting inaccuracy around the figures reported in the actual construction law, but it must have been clear for policy makers at the time that travel demand was a highly contested issue. Skamris (1994) found that

subsequent forecasts increased the expectations for demand substantially in the years after construction law was passed in 1987. However, a report from the ministry published two years prior to the construction law being passed indicates similar demand levels as the figures presented by Skamris (MOA 1985). Demand was expected to be 9,000-10,000 AADT in the construction law, while reports published both prior to and after the law being passed indicate demand in the range of 15,000-16,000 AADT. Actual traffic one year after opening was 22,000 AADT. It thus appears that the construction law was intentionally based on very conservative demand forecasts, which could explain why there is such a large positive inaccuracy for this project. It also appears that the estimates used for vehicle occupancy was much too high, since the travel volume is overestimated while traffic volume is underestimated. In addition to inaccurate demand forecasts, the project went 53% above the original budget, partly due to delays caused by fire and flooding³⁸.

The last project with traffic volumes above 200% of forecasts is a Danish motorway construction (Hadbjerg – Randers S), which was part of a national scheme to connect all larger Danish cities via motorways (typically referred to as the 'Big H'). The forecast was 6,500 AADT while the observed traffic was 18,100 AADT. However, the adjusted forecast is 8,800 AADT, since the original target year was 1985 but the project did not open until 1994³⁹. This improves the inaccuracy measure from 178% (raw) to 106% (adjusted) additional traffic, and while this is obviously a drastic reduction of inaccuracy it is still a quite dramatic deviation from the estimated figure. A general trend of demand underestimations can be witnessed for other projects completed in the East Jutland corridor at the time, indicating that demand in this region has generally been underestimated. Many of the forecasts for these projects were done by simply extrapolating older forecasts with the expected national growth factors at the time. However, travel demand in the region generally increased much faster than the national trends (VD 1995). In addition to inaccurate demand forecasts, the project went 32% above the initial budget, partly due to the need for an additional bridge and unforeseen geotechnical problems from tertiary clay deposits.

In the less extreme cases where there is still a considerable amount of additional traffic (between 50% and 100% above forecasts) we find a mix of bypasses, fixed links, and motorway constructions. Forecasting inaccuracy for some of them can be partly explained from arguments similar to that of the Hadbjerg – Randers project mentioned above, since they are part of the same national scheme and subject to the same erroneous assumptions for the regional demand expectations. Similar observations can be made for several UK projects in this span of forecasting inaccuracy, where regional demand has been underestimated by using national growth factors. An example of this is the Rushden – Higham Ferris project, where demand forecasts were based on expected national trends, but where local population growth was 14% compared to a 2% national average. Another possible explanation is the possible neglect of induced traffic in many of the demand forecasts, as

³⁸ While barely noticeable in the total budget, one of the more peculiar cost increases came from Denmark being sued by Finland at the International Court of Justice for blocking Finish-built offshore platforms, which resulted in an out-of-court settlement for around a hundred million DKK at the time.

³⁹ The estimated opening year was 1993 in the construction law, so this project is also a good example of why adjusted figures should be used rather than raw forecasts. No policy maker would have expected the forecast 1985 figures to pertain to a situation almost 10 years hence, and using these figures for evaluation of forecasting accuracy would be extremely misleading.

this often receives little to no attention in the available appraisal reports. It is very likely that this is a problem for most of these projects, as induced traffic has largely been ignored in transport models used in practise in all four countries at the time when these appraisals were made (Mackie 2010; MOTOS 2007; Nielsen and Fosgerau 2005; Welde and Odeck 2011). While induced traffic is important to address, failure to do so certainly cannot explain observed inaccuracies of more than 50% by itself for fairly standardized road projects, and there seem to be more fundamental problems in assessing input variables used to determine future travel demand as well as defining sufficiently large areas of influence for the models. However, it has not been possible to assess the impact of these erroneous assumptions and undertake a quantified assessment of how much of the observed forecasting inaccuracy they could potentially explain.

Projects that have more moderate amounts of additional traffic (between 20% and 50% above forecasts) display similar characteristics as the previous group, although a couple of motorway upgrades are also included here. Again we find some of the Danish projects from the East Jutland corridor, as well as several bypasses in the UK, where failure to assess regional development trends could be a possible explanation for the observed inaccuracy. There are also a couple of projects where it is fairly obvious that the expected traffic distribution has been a problem rather than overall demand. As an example is the Basford – Hough – Shavington project, where overall demand on the network was very accurately forecasted, but where the forecast failed to predict the distribution between the A500 and A534. This results in a measured forecasting inaccuracy of 37% for the actual project, but a forecasting inaccuracy of zero for the total network. Furthermore, the unexpected distribution of traffic can likely be explained with traffic calming initiatives near the A534, which was not known when the forecasts was prepared. This project thus highlights some of the methodological difficulties in assessing forecasting accuracy that I described in chapter 4, since this type of information is not available for many of the projects included in the sample of the present study. There might be very reasonable explanations why forecasters failed to predict the observed traffic volumes for individual projects, even when the deviation is quite dramatic.

In the group of projects that have a modest amount of forecasting inaccuracy (observed traffic volumes $\pm 20\%$ compared with forecasts) there is a larger share of newer projects than in the total sample and the vast majority of the motorway upgrades in the sample of the present study. It is difficult to make a qualitative assessment of common characteristics in this group, since it contains around two thirds of the total sample. Judging by the available decision support, most of the projects in this group are not expected to have considerable impacts on travel demand. However, this is a fairly trivial observation since we have already established that many of these projects are upgrades to existing links, where impacts are, *ceteris paribus*, easier to predict than for more drastic changes to the network. Perhaps the most interesting project in this category is the Øresund Bridge, where a forecast of 7,500 AADT in 2000 isn't far off the observed traffic of 8,300 AADT (no adjustment was necessary for this project). Considering the size of the project and the considerable network effects of a fixed link between Denmark and Sweden, the additional traffic does not appear to be a very problematic forecasting inaccuracy. However, in the original forecasts a static demand is listed for the years after opening, while road traffic volumes had already doubled seven years after opening, with rail traffic quadrupling during the same period. For a large project that is financed by user charges, the additional traffic is of course a welcome surprise when considering the payback period

of the project, but it also means that the Øresund Bridge might soon be operating at max capacity only 12 years after opening. This has led researchers and politicians in both Denmark and Sweden to voice the need of an additional connection to accommodate future demand, partly due to what is locally referred to as the ketchup effect⁴⁰ for migration and job markets (Berlingske 2008; Jyllands Posten 2011; Politiken 2007; Wessman 2006). On top of supply constraints across Øresund are of course also a greatly increased demand for local traffic in Copenhagen and Malmö, where capacity problems are already quite significant during peak hours.

In the category of projects that experience considerable less traffic than expected (between 20% and 50% less than forecasts) there is a considerable amount of bypass constructions. Given that bypass projects are typically among the projects with the highest underestimation rather than overestimation of traffic on average, this observation indicates that uncertainty is in general quite high for this type of projects. A closer look at some of the bypass projects in this group reveals that demand for the total network is typically more accurate, and that inaccuracy is often a result of failure to predict the actual distribution rather than the total demand of travel. Typical examples are Randers Ringboulevard in Denmark (a ring road) and the Alvaston Bypass in the United Kingdom. In the case of Randers, traffic on the new link was 36% below forecasts, while traffic on a nearby motorway was almost double that of the forecasts. In the case of Alvaston, traffic on the new link was 31% below forecasts, while traffic on the existing road was more than double that of forecasts. These projects further highlight the limitations of evaluating forecasting accuracy based solely on the new or upgraded links, since doing so completely ignores the possibility of total demand being underestimated in spite of severe negative bias on the new links. It is therefore regrettable that data for the overall network has not been available for all projects, as this makes it impossible to estimate whether such distribution issues blurs a general tendency of underestimating overall demand in cases where demand on the new link has been underestimated. Erroneous assumptions of key input variables that have likely led to a general overestimation of traffic can be observed for at least some of the projects in this group, although without the available data for the full network it is not possible to assess this in detail. Another important mechanism that the quantified inaccuracy measure does not address is the use of forecasts to display something else than the most likely scenario, such as a maximum estimate or a situation one wishes to avoid. As an example, the actual traffic for Björvikatunnelen under Oslo harbour was some 25% below the demand forecasts presented in the EIA for the project. However, in the same document it was explicitly mentioned that measures to ensure a lower growth should be taken, in order to end up with lower traffic volumes than those presented in the forecast and thus less environmental strain in the area. In this situation it would be wrong to consider the measured overestimation of traffic purely as forecasting inaccuracy, as policy makers were specifically advised to launch initiatives that would lead to the observed deviation between forecast and actual traffic volumes. The forecast itself was thus likely a contributing factor to the observed deviation.

For rail projects it appears that projects attracting more traffic than expected are typically relatively new and located in dense urban areas. As an example, Citytunnelen in Malmö had 49% more

⁴⁰ An expression used to describe cases where no immediate effect of an intervention can be observed, but where development suddenly changes dramatically, often to the surprise of observers.

ridership in the opening year than expected in the appraisal by the Swedish Transport Administration⁴¹. This large inaccuracy is perhaps not that surprising when we recall the ketchup effect mentioned earlier in relation to the Øresund Bridge. The demand forecast for Citytunnelen was prepared when the Øresund Bridge had just opened, but since then the travel demand between Denmark and Sweden has exploded. Since Citytunnelen is primarily linking Malmö's central transit hub directly with the Øresund connection, it is of course highly influenced by this massive increase in commuting between Copenhagen and Malmö. In addition, travel time between the two cities was reduced by 10 minutes on average, which is twice that of the forecast. Furthermore, the project was some 6% cheaper than expected, and using the updated figures for demand and costs reveals that the BCR of the project is 0.1 rather than the expected -0.46 (TV 2011). Another example is the first section of Bybanen in Bergen (light rail), where ridership in the opening year 2011 was already 7% above forecasts for 2015. SKYSS, the local transit authority, estimate that actual ridership is considerably higher than the official figures, since a lot of people with season tickets do not check themselves in on each ride. In addition, on-board ticket purchases are not counted either (TRM 2007). Bybanen was awarded 'Worldwide Project of the Year' in its opening year (TUT 2011), and is intended to serve as the backbone of transit services in Bergen. Part of the success seems to be a well-coordinated feeder system by rearranging existing bus services, which was also highlighted in the decision support as a key necessity for ridership forecasts to be accurate (SINTEF 2002). However, it is unclear if the project has actually managed to attract car users as stated in the forecasts⁴², or whether the increased ridership is mainly a result of former bus commuters, non-motorized commuters, or new trips entirely. From a feasibility perspective it is of course desirable that the increased ridership creates increased revenue, but from a planning perspective it is highly desirable to investigate whether the success of transit projects result in any congestion relief.

The rail projects with the most accurate forecasts are typically all upgrades to existing systems, such as additional tracks, new stations, or increased capacity at existing stations. Perhaps the most interesting project in this group is the Great Belt Bridge, where observed traffic was within 1% of forecast volumes. I have already discussed this project earlier when describing the different groups of road projects and shall not go into further detail with it here. A more typical example for this group would be the new double track on Frederikssundbanen that opened in 2002. Traffic was generally 7% lower than forecast for the overall line, but a station breakdown reveals that much of this can be explained by demand shortfalls on Kildedal station, where traffic was 88% below forecast. In the decision support, Kildedal was described as a future business hub with associated residential areas, and two years after opening a park and ride facility was constructed to attract further passengers (DSB 2004). However, the urban development never took place, ridership remains far below forecast for this station, and service shutdowns at night and on Sundays have since been implemented in a response to the lack of passengers. If we exclude this station from the assessment of forecasting accuracy for Frederikssundbanen we find that traffic is 3% higher than forecast on the rest of the line. The urban development plans around Kildedal were highlighted in the decision

⁴¹ A forecast prepared by Intraplan predicted much higher demand that was more in line with observed figures.

⁴² To the best of my knowledge there has been no study of travel behaviour that measures former car users for Bybanen in Bergen. COWI (2007) mentions that the project has managed to attract a lot of new transit users, but the data presented therein does not allow any conclusion as to potential modal shifts.

support as a key necessity for ridership forecasts to be accurate, and the project thus serves to illustrate both the uncertainty of input variables as well as the need for more than just aggregate ridership data in the assessment of forecasting accuracy. For stations on Frederikssundbanen and other rail upgrade projects, where no considerable changes in demand are expected, forecast accuracy remains very satisfying.

The rail projects that attract considerably less ridership than forecast are typically new inter-city lines and include both new and old projects. However, the airport express trains also appear in this group, with Arlanda Express as an example of a project with significant ridership shortfalls. Observed ridership is 65% below forecast, and it appears that certain assumptions in the decision support have been extremely optimistic. For example, standard tickets for Arlanda Express ended up at 360% of expected price levels, making the train much more expensive than competing express busses. Ironically, forecasts for Flytoget airport express train (Gardermoen – Oslo) were based on a model that was calibrated on data for commute patterns in Arlanda, but where only commute patterns for the most transit orientated commuter groups were included (VGTA 1992). In addition, the possibility of competing bus services was ignored in the forecasts⁴³. Forecasts for the total public transport share ended up being fairly accurate, but ridership on Flytoget ended up 51% below forecasts since bus services seized a substantial part of the market share. However, more typical projects for this group would be Swedish and Norwegian inter-city rail projects such as Jærbanen (Sandnes – Stavanger) or Mälärbanan (Stockholm - Örebro). The former is an upgrade from a single to a double track link where observed ridership is 27% lower than forecast, although this is still within the uncertainty span of $\pm 30\%$ that was listed in the decision support. The latter is a considerable overhaul of the transit services between larger cities where traffic was 61% below forecast, which is likely caused by optimistic assumptions regarding the autonomous growth in transit demand (BV 2006). In addition, it seems that quite severe errors from erroneous input variables were introduced in the transport model used to assess Mälärbanan, both in relation to ex-ante ridership figures and the planned time table. Common to most of the projects in this group seem to be that they are expected to attract ridership from competition with road-based alternatives (car or bus) on longer commute links, but fail to do so because assumptions regarding ticket prices, travel time, frequency, or connection options to other transit services has been too optimistic. These findings are also in line with possible bias in transport modelling techniques, where ridership in dense urban environments is likely to be underestimated, while ridership in less transit friendly environments is likely to be overestimated (Goetzke 2003).

6.3 The sensitivity of impact assessments

In this section I will briefly cover a set of critical cases to explore the impacts that forecasting inaccuracy has on subsequent impact assessments. For costs the impacts are trivial; cost overruns mean that a project has higher costs than expected, and thus, *ceteris paribus*, a worse investment than indicated in the decision support available to policy makers. For some projects there is a similar relationship between feasibility and forecasting accuracy of demand when the deviation is modest.

⁴³ Flytoget was built in connection with the new airport in Gardermoen. Gardermoen is now the central international hub for air travel in Norway, but prior to this it was only a reserve and charter airport. The decision to ignore bus alternatives thus seems to be based on an assumption that no bus services would be established rather than Flytoget outcompeting existing lines.

More traffic than expected typically increases overall monetary benefits for both road and rail projects, since a major budget item in the benefit side is travel time savings. This item is often perfectly correlated with demand, and for rail we typically have producer surplus as another item that is perfectly correlated with demand. However, demand is only a proxy for total project benefits, and the correlation between the two is not always linear or unidirectional. This can happen either when the economic surplus is not the most important motivation for the project, or when deviations in demand affect the benefit that each user experiences. For example, if a bypass is intended to move a lot of heavy through traffic out of a small town, then demand shortfalls might not be of vital importance to the overall benefits of the project, as long as the heavy through traffic has moved to the new link. Conversely, in heavily congested areas we might find that a new motorway generates more new traffic than expected, which reduces travel speed in the corridor due to congestion. For more extreme underestimations of traffic it might become necessary to expand capacity of fixed links earlier than expected, which is a lot more costly than building the link with this extra capacity in the first place.

As mentioned in chapter 4 I decided to focus this part of the analysis on a critical road case, since results for forecasting accuracy seemed more robust for road than for rail projects. I think the findings presented in section 6.1 support this interpretation, but I shall nonetheless spend some time on discussing a rail case in further detail. There are two reasons for this. First, while observed bias for rail projects is still less than for road projects, it is clear that there is still a significant trend of overestimating demand. It therefore seems relevant to address the potential implications of this trend for decision support. Second, one of the newer Danish rail projects (Ringbanen) had a considerable amount of ridership shortfalls, and data availability thus allows a more thorough analysis of this particular case. Since I do not have access to the original assessment tools I cannot replicate the assessment based on actual figures, but it is still possible to provide an informed analysis of how the ridership shortfall would likely have affected decision support.

6.3.1 Ringbanen

Ringbanen (The Ring Line) is a semi-circular train line that opened in the summer of 2007 and partially orbits Copenhagen. It is part of the local S-train transit system, and connects to many of the other transit services in Copenhagen, including the central bus station at Nørrebro and the start of the metro system at Flintholm. For most of the 20th century the corridor has not had a direct passenger rail service, but part of it was used actively for rail-based freight traffic until the 1990s. However, the Øresund Bridge resulted in freight traffic moving elsewhere, and it was decided to establish a ring line that would provide commuters with a direct rail link between Hellerup (N of Copenhagen) and Vanløse (SW of Copenhagen) without having to pass through central Copenhagen. In addition, it would improve linked trips for suburban commuters by offering a rail connection between the major radial corridors leading into Copenhagen, where commuters had previously been forced to transfer via bus services or continue downtown to connect with other S-trains. The alignment can be seen in Figure 40.

The original design has only received minor changes and the time schedule in use is the same as the one that was used in the preparation of ridership forecasts for decision support documents. Furthermore, the project finished on schedule, and existing transit options between Hellerup and



Figure 40: The suggested alignment for Ringbanen prior to passing the construction law. Apart from moving and renaming some of the stations slightly, no changes have been made to the overall design of the project. Reprint from Pilegaard and Johnsen (1999), translated by the author.

Vanløse provided a fairly good understanding of commuting patterns in the corridor. Many of the typical causes of extreme forecasting inaccuracy such as design changes, construction delays, and reduced frequency therefore do not apply to the Ringbanen case. However, ridership was about one third of the expected figures one year after opening, and construction costs were 35% higher than the initial budget. How did planners end up producing so inaccurate demand forecasts, and would the project have been scrapped if the actual ridership had been known at the time? These are some of the questions I will address in this section.

If we start with how the ridership forecasts ended up so inaccurate, the first place to look would be the original CBA (TRM 1998). The demand forecasts presented herein were used in both the subsequent EIA (BS 2000) and construction law (FT 2000), as well as referenced in press releases, news clips and conference presentations all the way up until the project officially opened in 2007 (e.g. BD 2004; Pilegaard and Johnsen 1999; Welin 2004). It is the same figures that are reported in all documents, and there can thus be no doubt about what forecast was used to inform decision makers or taxpayers about the expected ridership potential of the project. The original CBA describes the project as a substantial improvement to the transit options for people living in the suburban areas of Copenhagen. It is mentioned in the CBA that the most crucial elements of the quantitative analysis have been detailed assessments of the forecasts for construction costs and ridership potential, which fits well with the importance of demand forecasts outlined in chapter 1.

In the CBA we find that a variety of forecasts have been performed for conceptually different alternatives such as S-train, metro, and light rail in the corridor. In addition, two different models (HTM and OTM) were used to produce demand forecasts separately. Both models covered the greater Copenhagen region, but one (OTM) was far more sophisticated than the other (HTM). OTM was originally designed to assess public transport solutions when plans were still being made for the Ørestad region, but since then it has been updated and is now the standard transport model used for assessments in the greater Copenhagen area. Since the model was still quite new when assessing Ringbanen, HTM was used as a backup model to ensure robustness of forecasting results. An interesting observation is that HTM predicts 41% less ridership than OTM (TRM 1998), but apart from presenting the HTM results for comparison they are not referenced in the remainder of the report. The reasoning for this seems to be that the OTM model structure is better at describing the present situation and that the difference in ridership is believed to have no decisive influence on the economic appraisal.

Based on the appraisal figures presented in TRM (1998) we see that more than 80% of the expected annual benefits from the project are from reductions in travel time or fare revenues, both of which are directly correlated with ridership. The suggested solution had a 7.1% internal rate of return and a 50 year net present value of DKK 342 million. A sensitivity analysis for $\pm 20\%$ ridership is included, which shows a 4.6% and 9.4% internal rate of return and a net present value of DKK -60 and 745 million respectively. This clearly shows that the economic feasibility of the project is highly dependent on the accuracy of demand forecasts, and it should be obvious that a ridership shortfall of two thirds would have made the project appear much less favourable in an economic appraisal. Of all the appraised solutions it is only the metro design that comes out with a positive net present value in the -20% ridership scenario (DKK 311 million). However, additional economic appraisal of solutions that involve upgrades to existing S-trains rather than acquisition of new trains close much of the gap between this and the metro solution. Combined with the ability to act as a backup system for the existing S-trains as well as continued goods transport, S-trains ended up being the favoured solution in the final policy recommendation.

Since the actual ridership shortfall is much greater than the lower boundary of the sensitivity analysis, it seems safe to conclude that total benefit shortfalls are also quite severe. However, does this mean that the project is a failure? Judging from various media coverage and appraisal reports on

the Copenhagen transit systems, Ringbanen is generally communicated as a quite successful project (e.g. COWI 2008; DR 2007; DSB 2008; RH 2009) in spite of ridership figures far below those presented at the time of political approval. This is partly because the expected increase in ridership in the decision support documents was very high, with some stations being expected to serve four times the existing amount of passengers upon completion of the project⁴⁴. Such drastic increases in ridership did not occur, but Ringbanen has still managed to attract new ridership compared to the figures before the project opened. In addition to this, the project has made it both faster and easier for transit users to cross the metropolitan region, with headway of 5 minutes during daytime. It is already considered a vital part of the S-train network. It is also possible that some of the ridership shortfall is due to an inaccurate distribution of traffic flow rather than an actual demand shortfall, but this has not been investigated for the present analysis. Regarding costs, the majority of cost overruns for Ringbanen appear to be caused by additional project items that have been approved individually after the initial budget approval (TRM 2007).

Still, policy makers might have abandoned the project if the true ridership figures had been known prior to political approval, since the cost benefit analysis of the project would then have been very poor. However, if that is really the case, we ought to consider the validity of cost benefit analysis as an expression of societal benefits from transit investments. If service providers, tax payers, and policy makers alike appear to consider a project successful in spite of severe ridership shortfalls, a negative net present value, and a low internal rate of return, how useful is the CBA appraisal methodology then at expressing the value of a given project? Perhaps part of the problem lies in the reductionistic approach to benefit appraisal currently employed in CBAs for transit projects. The monetary benefits are primarily centred on consumers (time savings) and producers (fare revenue), but this approach ignores a lot additional benefits from transit investments. Banister and Thurstain-Goodwin (2011) has attempted to quantify non-transport benefits of rail projects at the micro, meso and macro level, and identified substantial additional benefits related to agglomeration economies, labour markets, environmental consequences, network economies, and property values. According to the authors, these effects are becoming increasingly important in areas that already have good transit coverage, which is a description that fits the Copenhagen case quite well. For some projects the inclusion of such additional benefits is reported to quadruple the benefit cost ratio compared to a traditional CBA (Banister and Thurstain-Goodwin 2011).

In this light it is possible that the philosophy of utilitarianism that CBA methodology is based upon is not necessarily underestimating benefits of transit projects due to its inability to reflect stakeholder preferences, but rather because the proxies we use for estimating the utility value are inadequate. Nonetheless, there are many aspects of society that are important to policy makers, but where CBA methodology is unlikely to ever express stakeholder preferences meaningfully. Examples of such could be distributional effects, intergenerational justice, or climate change. However, even if CBA methodology will never be able to cover such preferences we should at least expect it to be a

⁴⁴ Ringbanen is not expected to be solely responsible for this increase in ridership, as the Copenhagen metro was planned but not yet completed at the point of preparing these forecasts. The forecasts for Ringbanen are thus based on the combined effect of these two projects. However, the metro had already been completed for a few years when Ringbanen opened, and while there was some ridership shortfall compared to forecasts for this project, it was not remotely close to the levels of Ringbanen.



Figure 41: Alignment used for the Nordhavnsvej case (Tetraplan 2011). The proposed project offers a direct link between a busy motorway in the north-west and a new residential area planned at the harbour area to the south-east.

reasonable indicator of the economic value of a given project, but as the Ringbanen case shows this might not be the case at all under the current approach to appraisal.

6.3.2 Nordhavnsvej

The second case in this section is a simulation study, where two different models are used to forecast the travel demand for an urban road project. One model will ignore induced demand effects while the other will include short-term induced demand effects. The purpose is to compare the effect of the two model results on subsequent impact assessments. For a case study I have selected Nordhavnsvej, which is the largest municipal road project in Copenhagen in the last 30 years, and intended as congestion relief for the local road network in the dense residential areas nearby. The alignment used for the present case can be seen in Figure 41, which is the first section of the planned project. The entire project includes an underwater tunnel extension that connects directly to the Nordhavnen peninsula in the south-west corner of the map and is expected to be completed in 2015. Plans for roughly 800,000 square meters of residential development in Nordhavnen over the coming years are in place and Nordhavnsvej will connect this new urban area with a busy motorway to the west. The results presented in chapter 3 as well as section 6.1 clearly indicate a tendency of forecasts of demand for road projects to be underestimated, and particularly so for projects that involve the construction of entirely new links. As specified in chapter 4, this was the reason for including a critical case of the potentially disproportionate impact that forecasting inaccuracy of demand can have on subsequent impact appraisals in the form of CBAs or EIAs.

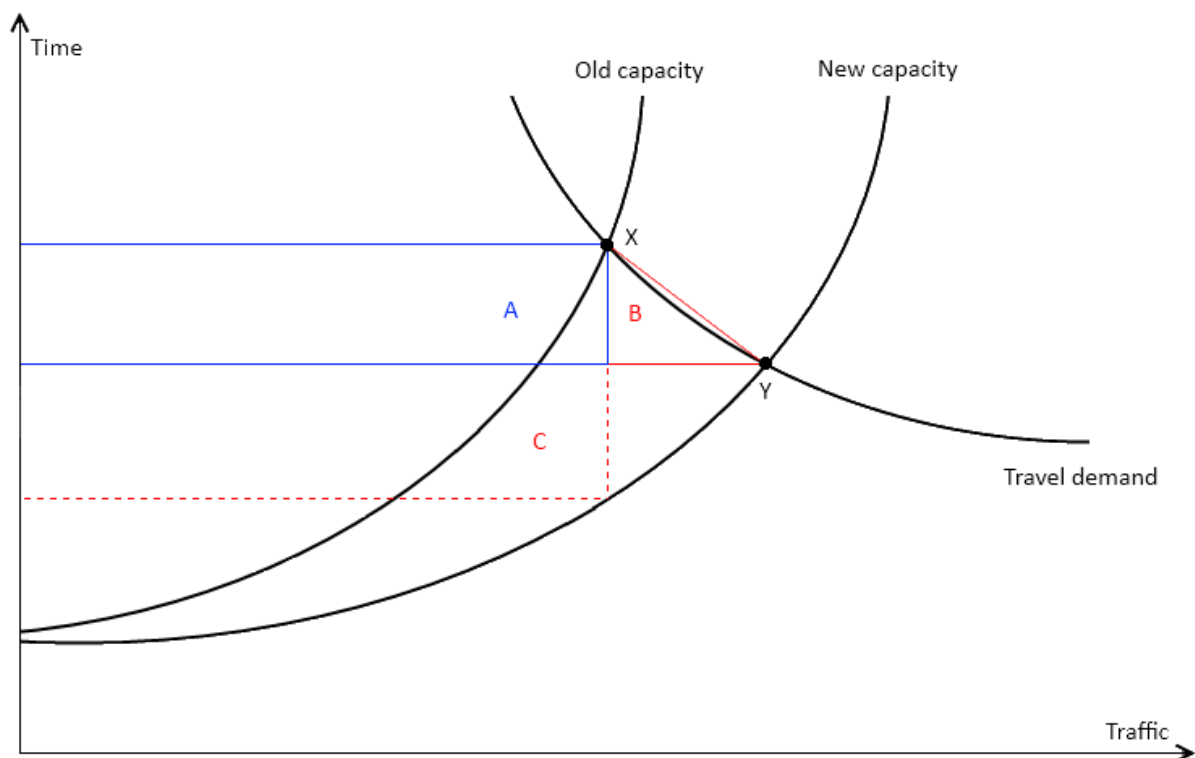


Figure 42: Illustration of induced travel demand by car as a response to faster travel times when new capacity is added to a congested network. Boxes A to C indicate the different estimates of travel time savings in cost benefit analysis. Nodes X and Y indicate the equilibrium states in scenarios with and without new capacity for current levels of demand.

Since population in the area is expected to increase dramatically over the next decade and capacity constraints are already problematic on the existing network, the Nordhavnsvej case is ideal for the purpose of illustrating the effects of induced traffic on impact assessments of road projects. In a scenario like this there is a latent demand for car travel since a lot of potential travellers suppress their travel activity as a result of congestion. They do so because the cost of engaging in such activity is higher than their willingness to commit resources to it. The resource requirements of travel are typically expressed as the generalized cost of travel, which does not only include the direct monetary cost of transport. The definition of these generalized costs varies, but typically they will at least include some measure of travel time in addition to the monetary costs. When the population increases in an area, the amount of potential travellers obviously increases proportionally, and the latent demand for car travel thus also increases. The introduction of additional travel options in the area can be seen as a reduction of the generalized cost of travel. For example, a new road will increase overall network capacity, which means it can carry more vehicles before it gets congested. Of course there are multiple potential bottleneck problems that can reduce travellers' ability to actually utilize this additional capacity, but the overall argument here is hopefully fairly uncontroversial. Since the generalized cost of travel is reduced by introducing additional network capacity, more people will now be willing to engage in travel activities. This is what we consider a release of latent demand, which is typically referred to as induced traffic⁴⁵.

⁴⁵ Induced traffic can be split in many different components, and the use of the term is often somewhat inconsistent between researchers, practitioners, and policy makers. This creates some confusion about what

A more formal representation of the above description can be seen in Figure 42, where we consider a proposed capacity expansion of a road network. This is a classic Marshallian Scissors diagram to express the relationship between supply and demand of a good, which in this case is the supply of network capacity and the consumption of it by users. The price of our commodity is travel time, but it could in theory take on any definition of generalized travel cost. The illustration mainly serves as a pedagogical reference, and in reality people's travel behaviour is obviously more complicated and dependent on many other aspects than travel time, just as the relationship between supply and demand is not at all as trivial as presented here. However, it is a fairly accurate representation of how the concepts of travel time savings are evaluated in CBAs. There are two supply curves to illustrate the network capacity for both the do-nothing (zero-alternative) and do-something (add capacity) scenarios. In the case of congestion from too much traffic, network speed slows down and travel time increases. The supply curves thus change from horizontal to vertical. Without new capacity we notice that the supply curve changes to a vertical shape at a lower traffic volume than with new capacity. This is due to the smaller network's inability to serve the same traffic volumes before congestion occurs.

Given these two scenarios of supply and demand levels respectively, we observe two different equilibriums at the points X and Y in Figure 42. In the short-term we observe that additional capacity reduces travel times. This in turn attracts additional travellers, and the equilibrium moves from X to Y. The evaluation of benefits for this situation is quite simple in principle, but somewhat more problematic to assess accurately in practice. The aggregate travel time savings are simply described as the number of travellers that enjoy reduced travel times multiplied with the time savings enjoyed per traveller. A monetary value is then assigned to this aggregate figure based on standardized prices. It should be obvious that not all travellers enjoy the same amount of time savings or value it equally, but by using average figures and splitting travellers into separate groups it is possible to obtain a reasonable approximation of the aggregate benefits. An important distinction is made for new travellers, who are only considered to benefit half as much from reduced travel times as existing users. This is known as the 'rule of half' and indicates that these travellers enjoy only marginal benefits, which is illustrated by the triangle B in Figure 42. The total benefits of adding new capacity at present levels of demand is thus the area A + B (existing and new users respectively). However, if we fail to consider the induced demand released by the new capacity we end up estimating the benefits as A + C instead. This means that the benefits for new users is ignored (B), but we mistakenly assign an additional value to the benefit for existing users (C). I say mistakenly, since this estimate does not take into consideration the effect on congestion on travel time, and we thus end up assigning existing users shorter travel times than they are likely to experience. Since the area C is much larger than the area B we thus end up with a larger total benefit estimate when we ignore induced demand, and our CBA thus ends up portraying the new road scheme as more economically feasible than is likely to be the case. As the present case is intended to illustrate, this difference is not always trivial and can have a profound impact on policy-making.

effects are actually referred to when one speaks of induced demand. For the present scenario I define changes to destinations and additional trips as induced traffic, while the general growth trend and traffic redirected from other routes are included in both models.

As already mentioned, Nordhavnsvej is a case where we would expect a considerable amount of latent demand to exist due to present levels of congestion, and both the short and long-term scenarios described in Figure 42 are very likely to apply here. In order to illustrate the consequences of inadequate treatment of induced demand in ex-ante appraisal I shall present two different ex-ante appraisals for Nordhavnsvej. The first will be based on a transport model that does not take induced demand into account. The second will be based on a transport model that takes some of the short-term effects of induced traffic into account. In spite of criticism during the latter half of the 20th century, the theory of induced traffic is now generally accepted among transport planners (American Association of State Highway Officials 1957; Downs 1962; Goodwin 1996; Growth 1963; Hills 1996; Litman and Colman 2001; Mogridge 1990; Noland and Lem 2002; Næss, Mogridge, and Sandberg 2001; Overgaard 1966; SACTRA 1994; Thomson 1977). However, due to the complex interaction between transport options, land use, and travel behaviour it is difficult to isolate and quantify the effects of induced traffic from policy interventions. Existing methods for doing so are primarily based on elasticity values between the generalized costs of travel and the total travel demand, which is well in line with the utilitarian approach used to assess the benefits of travel time savings described earlier in the present section. However, since models can differ immensely in their internal structure, we will need to compare the results of appraisals that are fairly identical in the structural setup, apart from their treatment of induced demand. For the Nordhavnsvej case we have chosen to use the existing OTM model to represent an appraisal that takes induced demand into account. A modified version of the same model is used to represent an appraisal that does not take induced demand into account. This modified version will therefore only evaluate changes to route assignment, and not at changes to destinations, mode choice, or new trips.

In order to ensure that the results are as closely in line with standard practice as possible, the consultancy firm Tetraplan was hired to perform all calculations involved in the appraisals using the current transport economic unit prices defined by the Danish Ministry of Transport. The decision to adopt this approach also allowed Tetraplan to verify that the modified version of the OTM model was a reasonable representation of the simpler models that are often used in practice outside the Copenhagen region. The appraisal does not take into account long-term induced demand such as residential relocation, increased car ownership, or deterioration of transit options. Furthermore, growth in demand after opening is assumed to roughly 10% over a 12 year period, after which a static demand situation is expected. The overall demand can thus be considered quite conservative and applying only to the short-term effects immediately after project opening. For further details about the assumptions for the present modelling procedures I refer to the summary report by Tetraplan (2011) and for details about the OTM model I refer to Jovicic and Hansen (2001). The results from the appraisal as well as a discussion of their implications to project appraisal has also been used in other publications from the UNITE project (e.g. Næss, Nicolaisen, and Strand 2012).

A comparison of the results of the two appraisals can be seen in Table 11, where the model excluding induced demand is labelled A and the one including it is labelled B. It should be noted that certain effects that are typically included in a CBA have been excluded for this analysis, since they are not expected to have any significant influence on the result. This is primarily done to reduce resource use devoted to external consultancy. The items that have been excluded are mostly related to costs,

#	Impact factor	Model A	Model B	$\Delta A \rightarrow B$
1	Traffic on main link (AADT)	21,740	22,820	+1,080
2	Travel time savings (mil. DKK)	4,589	2,749	-1,840
3	Changes in fuel consumption (tons)	-284	+483	+767
4	Changes in CO ₂ emissions (tons)	-897	+1,525	+2,422
5	Changes in noise level (weighted score ⁴⁶)	162	+167	+5
6	Changes in safety (accidents involving injury)	-1.2	-0.3	+0.9
7	Net present value (mil. DKK)	2,157	403	-1,754
8	Internal rate of return (%)	8.1	5.6	-2.5
9	Benefit ratio per invested capital unit	1.1	0.2	-0.9

Table 11: Results of a cost benefit analysis of the Nordhavnsvej case based on two different transport models. Model A does not take induced demand into account while model B accounts for short-term induced demand. The last column indicates the change in appraisal figures for individual items as a result of including induced demand in the model.

since this was not a focus area for the case study but obviously still required to calculate the last three items in Table 11. Construction costs have thus been based on cost estimates from existing impact assessment reports, while nuisances during construction are not included. These effects would likely have very limited influence on the total costs of the project, just as all such cost related items would affect both appraisals equally. The costs are thus slightly underestimated, and care should be taken when interpreting the appraisal figures for the last three items in Table 11 as they are likely inflated as a result of this. It is the comparison of the two appraisals that is of chief interest here, and the observed difference between them is thus more important than the absolute figures.

From Table 11 we observe that expected traffic volumes on the main link increase slightly by including induced demand in the transport model (Table 11, #1). This additional traffic corresponds to around 5% inaccuracy. Of course this figure is predicted rather than estimated, and many uncertainties remain that could still cause the figure for model B to be very inaccurate as well. However, for the present case we are mainly interested in the relative difference between two appraisals, and we observe that this is well in line with the general trend in the empirical observations of forecasting inaccuracy presented throughout the present thesis. We are thus able to compare the consequences of additional traffic on impact assessments for a critical case.

⁴⁶ The weighted score is expressed as SBT (støjbelastningstal), which is a standard unit in Danish noise evaluations. A higher score indicates a higher noise level, although in this case the difference is fairly insignificant.

One immediate observation is that the monetary benefits from travel time savings are much lower when including induced demand (Table 11, #2). The 5% additional traffic apparently results in 40% less benefits. We might recall from earlier chapters that travel time savings constitute the vast majority of total monetary benefits in CBAs, so this is quite a drastic figure. Failure to account for short-term induced demand would result in an overestimation of consumer benefits of almost 2,000 million DKK in this case. Note that the future growth in demand was considered to be very conservative, and that this comparison only accounts for short-term induced demand. If demand increases more rapidly than this conservative estimate, there might be an even further loss in travel time savings due to congestion.

If we look at the items from the environmental assessment we observe that the inclusion of induced demand increases the expected impact of each major category (Table 11, #3-6). I have included these in their impact units rather than converted them into monetary figures. There are two reasons for this. First, the pricing of unit values for environmental factors seems to be disputed much more than the pricing of unit volumes for travel time savings (see e.g. Van Wee 2011). For example, some would question the rationality of comparing non-commensurable categories such as environmental harm and faster travel times directly, since we cannot use the time saved to buy back the environment destroyed in this direct manner. Pricing is typically based on willingness-to-pay studies or quota systems, but these prices reflect market demand rather than the actual cost of reversing the process if we were to undo the damage at later stages. Some categories of environmental harm may even be irreversible⁴⁷, of which climate change would be a good example. Second, for reasons similar to those above, the discounting of environmental harm is also disputed. Discounting of future costs and benefits is based on the rationality that income is rising and that one unit of something now is thus worth more than it would be three years hence. This is true in-so-far that income is actually rising and that the costs and benefits are evaluated similarly in the future. However, since willingness-to-pay is used to price many environmental factors, it is likely that people's willingness-to-pay for environmental protection might also increase as their income levels increase. In this case we should use a negative discounting rate for such items. This is especially true for irreversible harm, if the environmental damage reaches some sort of critical threshold level. These things are completely ignored in current CBA methodology. Third, studies have shown that decision makers do take environmental factors into account when evaluating costs and benefits of a project, but they typically do so in a non-monetized form. Odeck (2010) finds this to be the case for Norwegian road projects, and speculates that it is likely due to controversies regarding the pricing of unit values. Policy makers thus seem to prefer non-monetized units as it can otherwise be difficult to evaluate the actual consequences of their decisions. Similar arguments can of course be made for travel time savings, but this seems to be a more institutionalized item for monetary evaluation than environmental impacts. Nonetheless, policy makers are unlikely to evaluate only the aggregate monetary benefits of travel time savings.

⁴⁷ This is true at least within timeframes of 100 or 1000 years depending on the impact category. Whether these impacts are actually reversible beyond such timeframes seem irrelevant to policy-making, since the damage would still last for generations. This probably relates to a more complex ethical discussion of intergenerational justice that is beyond the scope of the present case.

Travel time savings (hours)		Model A	Model B	$\Delta A \rightarrow B$
Cars	Existing drivers	1,282	834	-448
	New drivers	0	37	37
Trucks	Existing drivers	284	207	-77
	New drivers	0	6	6
Travel time savings in congestion (hours)		Model A	Model B	$\Delta A \rightarrow B$
Cars	Existing drivers	874	435	-439
	New drivers	0	16	16
Trucks	Existing drivers	163	85	-78
	New drivers	0	1	1

Table 12: Travel time savings for the Nordhavnsvej case expressed in non-monetary units based on two different transport models. Model A does not take induced demand into account while model B accounts for short-term induced demand. The last column indicates the change in appraisal figures for individual items as a result of including induced demand in the transport model.

The aggregate socioeconomic indicators show that inclusion of induced demand reduces the perceived viability of the project quite drastically (Table 11, #7-9). Net present value for model B drops to less than one fifth of that in model A, primarily due to the reduced benefits from travel time savings. The internal rate of return drops below the 6% recommended for public investments and the return on invested capital is also much lower. Considering the large difference in travel time savings (Table 11, #2) it is fairly unsurprising that the overall performance of the project suffers. However, what might be surprising is that a slight increase in traffic on the main link has such dramatic impacts on the subsequent impact assessments. If we dig a little deeper in the data for travel time savings we will see that this is indeed due to an overestimation of time savings per user in model A.

Table 12 displays a breakdown of the travel time savings expressed in non-monetary units. It should be obvious that there are no benefits assigned for new drivers in model A, since we do not include induced demand in this model. The interesting part for the present case is the comparison of benefit losses for existing drivers with benefits gains for new drivers when moving from model A to model B. Here we clearly see that the additional benefits enjoyed by new drivers (area B in Figure 42) are much smaller than the loss of benefits to existing drivers (area C in Figure 42). This is especially the case during peak hours, and we see that the ability of a new road to serve as congestion relief is severely overestimated. Existing drivers are expected to enjoy twice the amount of congestion relief in model A compared to model B. Since congestion relief is often a motivation for expanding road

capacity in urban areas, this is a severe distortion of the scheme's ability to achieve the desired policy goals.

6.4 Summary

The purpose of this chapter has been to present the first part of the analysis related to how well expectations in ex-ante appraisal compare to the actual development ex-post. The accuracy of demand forecasts has been the primary unit of analysis in this regard, which was split on road and rail projects. For road projects, the observed inaccuracy for demand forecasts is well in line with most previous studies, but splitting the road projects into subgroups of different types revealed that not all road projects are equal in terms of bias. For rail projects, the observed inaccuracy for demand forecasts is much smaller than in most previous studies, albeit still displaying a tendency of overestimation. Cost estimates were generally fairly accurate for both road and rail projects compared to previous studies. The chapter ended with two examples of how difficult it can be to assess benefits in transport project evaluation, and that the accuracy of demand forecasts are not a very good proxy for evaluation of project benefits.

7 Uncertainty management

It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.

- Mark Twain

In this chapter I present what I consider the most important findings in the study of how uncertainty is communicated and handled among different stakeholders. The analysis involves both the responses to the questionnaire sent out to stakeholders in Scandinavia and the information obtained during qualitative interviews with a select few stakeholders at different positions in the decision-making hierarchy. This chapter is more explorative in nature than the previous, and especially so for the analysis of interviews. The purpose of the chapter is primarily to gain some insight into what mechanisms that might cause the observed inaccuracies that were reported in the previous chapter, as well as how stakeholders go about planning for the future in the face of uncertainty. Much of the latter point concerns the communication of uncertainty, which is a central theme throughout the chapter.

7.1 Uncertainty in decision support documents

The second research question (cf. section 1.3) addresses how uncertainty is managed in decision-making processes. As described in chapter 4, the available decision support documents for the analysis of the first research question would also be well suited to study the communication of uncertainty in the decision-making process. There are several reasons for this, of which the most obvious is perhaps that these documents are supposed to inform policy makers of the likely consequences of their decisions. This naturally includes the inherent uncertainty that is associated with making predictions about the future development in complex open systems.

As described in section 5.1 it has not been possible to obtain the original decision support documents for all projects. Consequently, it has not been possible to analyse the communication of uncertainty in the projects where only forecasts are available but not the context they are presented in. The sample is thus smaller for the present analysis than the analysis in presented in section 6.1 that draws on comparable sources of evidence. However, there are no apparent reasons for why the sub-sample used in the present analysis should lead to results that are not comparable with those from the total sample, as the exclusion of projects in this fashion must be considered fairly randomized⁴⁸. Given the available data, it would introduce much greater concern for the validity of the results if the communication of uncertainty was partly based on construction laws and ex-post auditing. These cannot be expected to fulfil the same role of decision support as the actual impact assessments prepared for policy makers, and the degree to which uncertainty has been communicated herein would thus be a poor indicator of the available information in the decision-making process.

⁴⁸ A quick analysis of these projects also revealed that the overall distribution of uncertainty is more or less identical to that of the full sample.

Rail projects have been particularly difficult in this regard, and will therefore be excluded from the quantitative analysis. These will be treated qualitatively, where relevant examples have already been introduced in section 6.2.5. The main focus will instead be on Danish and Norwegian road projects, since the availability of original decision support documents have been the best in these two case countries⁴⁹. It has been possible to analyse 78 projects regarding the communicated level of uncertainty related to the demand forecasts, and 76 projects regarding the communicated level of uncertainty related to the cost forecasts⁵⁰. Uncertainty regarding demand forecasts has been further broken down into the location dimension, while this has not been possible for costs due to reasons described in section 5.2. The level and location of uncertainty related to the concepts defined in section 2.3, and the level dimension will thus be split into four different categories:

0. Not mentioned: This level denotes decision support documents that do not address uncertainty at all, and therefore provides no explicit information for policy makers on the validity and reliability of the conclusions presented in it.
1. Risk: This level denotes decision support that includes quantified levels of uncertainty for the relevant appraisal indicators, typically in the form of a confidence interval, triple estimate, or spans rather than point estimates.
2. Doubt: This level denotes decision support where sources of uncertainty are dealt with more qualitatively and typically with causal direction for the relationship between these sources and the relevant appraisal indicators.
3. Ignorance: This level denotes decision support in which unspecified sources of uncertainty or unspecified direction of specified sources of uncertainty are described, and thus provides some context for the conclusions presented in it.

From the framework presented in section 2.3 the categories should be considered ordinal in the sense that a higher number indicates a higher degree of perceived indeterminism in the results communicated in decision support documents. Given the mathematical representation of transport models used for demand forecasts, it is of course desirable that the knowledge is as deterministic as possible. A lower score indicates that results are presented as highly deterministic, while a higher score indicates that results are associated with large uncertainty. In the event that no communication of uncertainty is included at all, this is considered a presentation of forecasts as completely deterministic. This is thus what Walker et al. (2003) labels as the far left of their uncertainty spectrum (see Figure 43), and is considered an unobtainable ideal that is not given a separate category in their framework. While it is true that this is an unobtainable ideal for transport forecasts as well, it is labelled as a separate category in the present analysis for the practical purpose of displaying how uncertainty is sometimes poorly communicated (or in such cases, not communicated at all) to policy makers, since they are presented as unrealistically deterministic. This

⁴⁹ For readers who missed it, the obvious reason for this was described in section 5.1. This is perhaps a more serious concern regarding the potential representativeness of the sub-sample used in the present analysis, if there is reason to believe that significant differences exist in the way that uncertainty is handled in decision support documents in Sweden or the UK compared to Denmark and Norway, I do not consider this to be a critical issue. Based on my personal experience while gathering the empirical data for the present thesis, this does not at all seem to be the case.

⁵⁰ Two projects had separate reporting of construction costs and the available decision support thus only made it possible to make valid claims on the communication of uncertainty in relation to the demand forecasts.

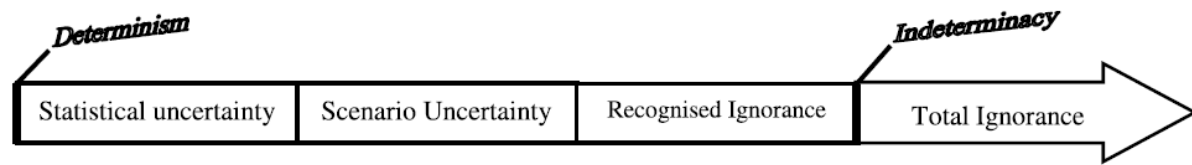


Figure 43: The progressive transition between determinism and total ignorance. Reprint from Walker et al. (2003).

also means that the scale is nominal rather than ordinal regarding the quality of uncertainty communication, since a low score could just as well be an indication of suppressed uncertainty rather than the quality of forecasts. Likewise, a high score says nothing about whether the communication of uncertainty has been useful. For example, stating that forecasts are inherently bound to be highly uncertain might communicate some degree of indeterminism, but such a statement is of little use to policy makers seeking to use the forecast as decision support. Rather, the scale should probably be treated nominally for this purpose and as a measure of how uncertainty is typically communicated in decision support. The implications of the results will need to be discussed more qualitatively. The nature dimension of uncertainty (cf. section 2.3) will not be given any attention in the present section, as the relatively spare description of uncertainty in the available decision support documentation provides no foundation for such analysis. It is likely the case that all instances of uncertainty presented here are a mix of both, as there is both a high degree of variability as well as problem of data availability when preparing demand forecasts as well as cost estimates. For policy makers it is likely also fairly uninteresting what the nature of a given uncertainty is, while the potential impact is of much higher concern.

7.1.1 Uncertainty in relation to demand forecasts

Figure 44 displays the uncertainty level for demand forecasts reported in decision support for road projects where the available quality of documentation fulfils the requirements outlined in the beginning of this section. It is evident from this illustration that the vast majority of documentation offers little or no information about the uncertainty of forecasts to policy makers. In fact, less than one fifth of the projects address the issue of uncertainty at all in the available decision support. In the few documents that provide a quantified estimate for uncertainty, it is typically in the range of $\pm 10\%$ for larger links and $\pm 20\%$ for medium links. As an example, in the environmental impact assessment for the expansion of Køge Bugt motorway it is stated that *"the uncertainty in the calculation of car traffic on the largest roads dominated by regional traffic is considerably lower [than any random stretch], typically around 10%"* (VD 2003, 83). However, as we witnessed quite clearly in the chapter 6, the uncertainty in forecasting travel demand can lead to much larger inaccuracies when compared with actual development, which can have considerable impact on subsequent impact assessments. It is true that various documents lists some of the more important assumptions that are used to produce the forecast, such as expectations for other infrastructure projects in the corridor or specific details on how the generalized transport costs are calculated. This is a reasonable argument in the sense that it communicates the conditions for the forecast to other modellers, who are then able to judge the validity and reliability of the results to some degree (depending on the detail of this information). However, it is unlikely that a given policy maker is able to infer the same conclusions based on this information. She might agree or disagree about whether it is reasonable to assume a

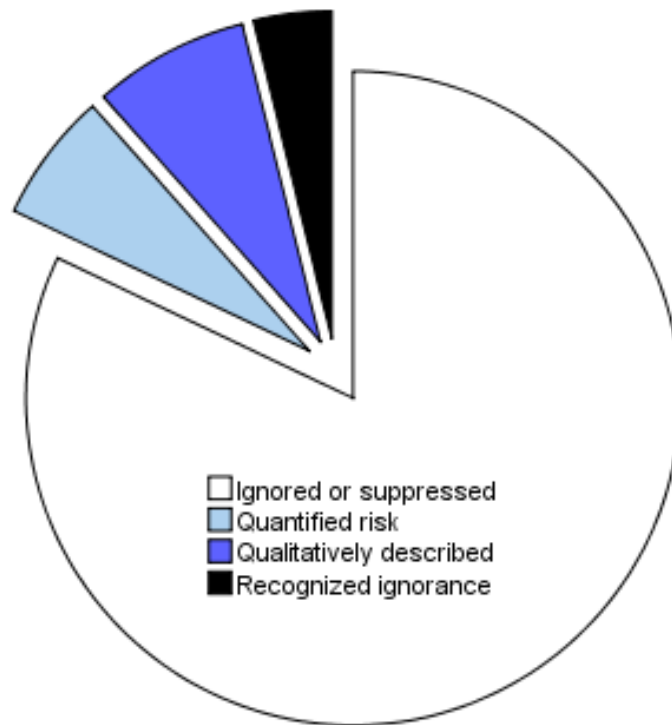


Figure 44: Uncertainty level for the demand forecasts as reported in decision support documents for road projects (N=78).

static peak hour spread under rising levels of congestion, but does she know what the impact of this assumption means for other parts of the system? She might also know what an all-or-nothing assignment model means conceptually, but does she know what the implication for appraisal will be in practice? It is highly speculative to presuppose that policy makers are able to understand the full implications of the assumptions embedded in model-specifications, and for this reason stated assumptions should not be considered a particularly valuable communication of uncertainty to policy makers. Even if policy makers were to make sense of the complex web of relationships between these assumptions and the appraisal results, these specifications are not particularly abundant in the available decision support either.

Among the relatively few documents that include a discussion of uncertainty in relation to demand forecasts, the uncertainty location is typically focused on the development of input variables that are external to the model (3 in 4 documents). This uncertainty is rarely presented as a quantified risk, but typically discussed qualitatively in terms of doubt. The model structure and its likely impacts on results are sometimes discussed explicitly (1 in 2 documents), and typically in the framework of recognized ignorance. Parameter and context related uncertainties are almost never brought up in decision support documents, but when they are it is typically also in the form of recognized ignorance. These auxiliary observations are presented in brief here since they have a strong relevance to the overall objective of the present analysis. It is therefore unfortunate that a vast majority of documents do not address uncertainty at, since there is then little purpose to a more nuanced analysis of how the location dimension is treated. However, the observation that

uncertainty is not given much attention is of course an important realization in itself, just as the focus on input related uncertainty in the places where it is actually discussed can be used as a point of departure for discussing if and how the concept of uncertainty needs to be more explicitly addressed in decision support for policy-making.

The observations listed in this section can be interpreted in a number of ways, of which a few that immediately come to mind are presented with some reflective comments. It seems that most people (experienced researchers being no exception) will attempt to explain such observations based on some form of abductive reasoning, or what Harman (1965) would likely label as inference to the best available explanation. These reflections are therefore intended as a pre-emptive response to some of the likely inferences that might be offered to the observations presented above.

First is an argument of context relevance. The demand estimates might not be considered of critical importance to the specific problem being addressed in cases where no uncertainty is reported. An example could be a junction improvement for a stretch of motorway, where the project is more focused on convenience for drivers than on providing capacity. While this seems a plausible explanation for some projects in the sample for the present study, it is certainly not the case for all of them. Most of the projects either attempt to divert traffic around residential areas, meet rising levels of demand, or improve travel time between two poorly connected regions. In all of these cases the demand forecasts are important in determining the expected efficiency of the project, since the project benefits are then either primarily dependent on the volume of diverted traffic or the volume of passengers benefiting from faster travel times⁵¹. This argument thus seems a poor rationale for excluding a discussion of uncertainty in the decision support documents. The projects that discuss uncertainty explicitly are among the larger, more complex projects though, where the quality of demand forecasts must be expected to have the biggest impact. This means that when uncertainty of demand forecasts is addressed, it is typically also very relevant to the project. It should thus probably be considered a necessary but insufficient criterion for proper communication of uncertainty, and not a support of the relevance argument that is being addressing here.

Second is an argument of impact severity. The level of uncertainty in the demand forecasts might be considered negligible for the validity and reliability of the forecast in decision support documents. This could be an issue in cases where demand has been relatively steady for a longer period of time and where no substantial changes are made to the system. However, the observations presented in chapter 6 indicate that uncertainty often results in quite large forecasting inaccuracies, so this can only be true for a limited part of the projects from the sample of the present study. Furthermore, even in cases where the demand forecast turned out to be accurate, this doesn't address any problems of uncertainty in relation to the zero-alternative, as this must be considered a counter-factual scenario. Nor does it account for cases where the accuracy is a result of coincidental circumstances other than those described in the decision support, but which led to a similar level of travel demand. This argument thus seems to be rooted mostly in ignorance towards uncertainty and the accuracy of demand forecasts.

⁵¹ The benefits from travel time savings also include the magnitude of time saved per passenger of course, which is related on the level of demand. This was discussed in further detail under the Nordhavn case in section 6.3.2.

Third is an argument of source selection. While the observations might indicate uncertainty to be handled poorly in the available decision support, this can be explained with complexity of demand forecasting being handled in technical sub-documents that address this issue in detail. The argument has some merit in the sense that it is customary for both CBA and EIA documents for larger projects to contain a summarizing final report as well as several sub-reports on specific topics being addressed. For transport infrastructure projects this typically involves a traffic impact study (TIS) that is reported in a separate document. However, these sub-documents do not to address uncertainty management in a suitable manner for the context of policy-making as discussed in the present study. Since the TIS are technical reports they are written for a specific audience, which can make it difficult for non-experts to understand the information presented within. Most policy makers and citizens can be categorized as non-experts when it comes to transport modelling. While the TIS might include more elaborate communication of the methodology, this is typically confined to a descriptive account of data collection, model choice, and calibration process. This is very useful information for other modellers, but not so much for laymen. It would thus be difficult for non-experts to use the information in the TIS as an assessment of the uncertainty related to the demand forecasts and the impact on subsequent impact assessment. A related source selection argument can be made regarding the informal settings in which much policy-making takes place. It is true that uncertainty can (and should) be a topic in such settings, which might counter some of the critique put forward here. However, while it is fruitful to use these informal settings as a forum for discussing uncertainty it cannot be considered satisfactory if it is confined to such forums. This makes auditing of such practice impossible and does not create equal awareness of uncertainty to the public. In fact, it might be used as a tool for informed policy makers to distort the decision-making process, if such information becomes privileged to a select few stakeholders, whether intentional or not. The source selection argument regarding informal communication channels thus faces a severe problem of transparency in a (supposedly) democratic system of governance.

7.1.2 Uncertainty in relation to cost estimates

As explained in chapter 1 and in the beginning of the present chapter, cost data has been secondary to demand data. The projects included in the present analysis are thus based largely on availability, and due to reasons highlighted in chapter 5 it has not been possible to obtain much data that would allow any insight to the location dimension of uncertainty. Thus, only the results of the level dimension are presented to compare with the results presented in section 7.1.1 for demand forecasts.

Figure 45 displays the uncertainty level for cost estimates reported in decision support for road projects where the available quality of documentation fulfils the requirements outlined in the beginning of this section. As was the case for demand forecasts we observe that uncertainty is not addressed in the decision support for a majority of projects, although this share is a little lower for cost estimates than demand forecasts. More interesting perhaps is the observation that cost estimates are typically quantified when they are addressed, whereas demand forecasts tended to be addressed at a variety of different uncertainty levels. This could be interpreted as construction costs having a higher priority than demand forecasts in terms of risk management, since cost overruns can be quite severe. It could also be interpreted as a cost estimates being easier to account for in detail than demand forecast, and thus be associated with a higher degree of determinism. It is difficult to

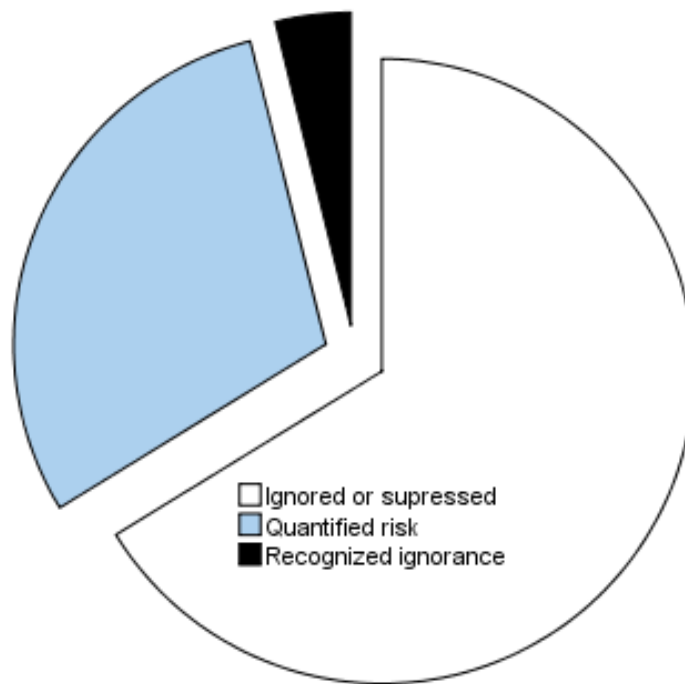


Figure 45: Uncertainty level for the cost estimates as reported in decision support documents for road projects (N=74)

assess any of these claims on the basis of the observed tendencies here though, as the process of quantifying this risk is not always described very well in the available documentation. As an example, in the decision support for most of the Norwegian road projects it is stated that the actual costs must fall within $\pm 25\%$ of the cost estimate, but how this risk level is ensured is not very clear from the documents themselves. There does appear to be some correlation between the reported uncertainty level and the inaccuracy of cost estimates presented in chapter 6, indicating that projects that do not report any uncertainty for cost estimates also tend to have lower levels of inaccuracy than those that do report uncertainty. However, this correlation is very weak and not at all statistically significant, perhaps due to the few projects for which cost estimates, final costs and reported uncertainty levels are available. Cost data for the mere comparison presented in chapter 6 was already sparse, and adding further data requirements reduces the available sample quite drastically. In addition, there is no indication that the observed inaccuracy corresponds to the quantified risks for a given problem. These observations could be interpreted as uncertainty being included for the projects where uncertainty is most likely to be an issue, but that the degree of reported uncertainty is unlikely the potential degree of cost overruns.

7.2 Survey results on uncertainty management

In this section an analysis of the responses to the questionnaire sent out to stakeholders in Scandinavia is presented. I refer to the earlier sections 4.4 and 5.3 for elaboration on the overall focus and target group of the questionnaire. In cases where an N value of less than 453 (total number of respondents) is used, this indicates that respondents answering 'don't know' have been excluded or the results presented account for only a specific group of respondents. In the latter case the respective subgroup will of course be made explicit when presenting the associated results.

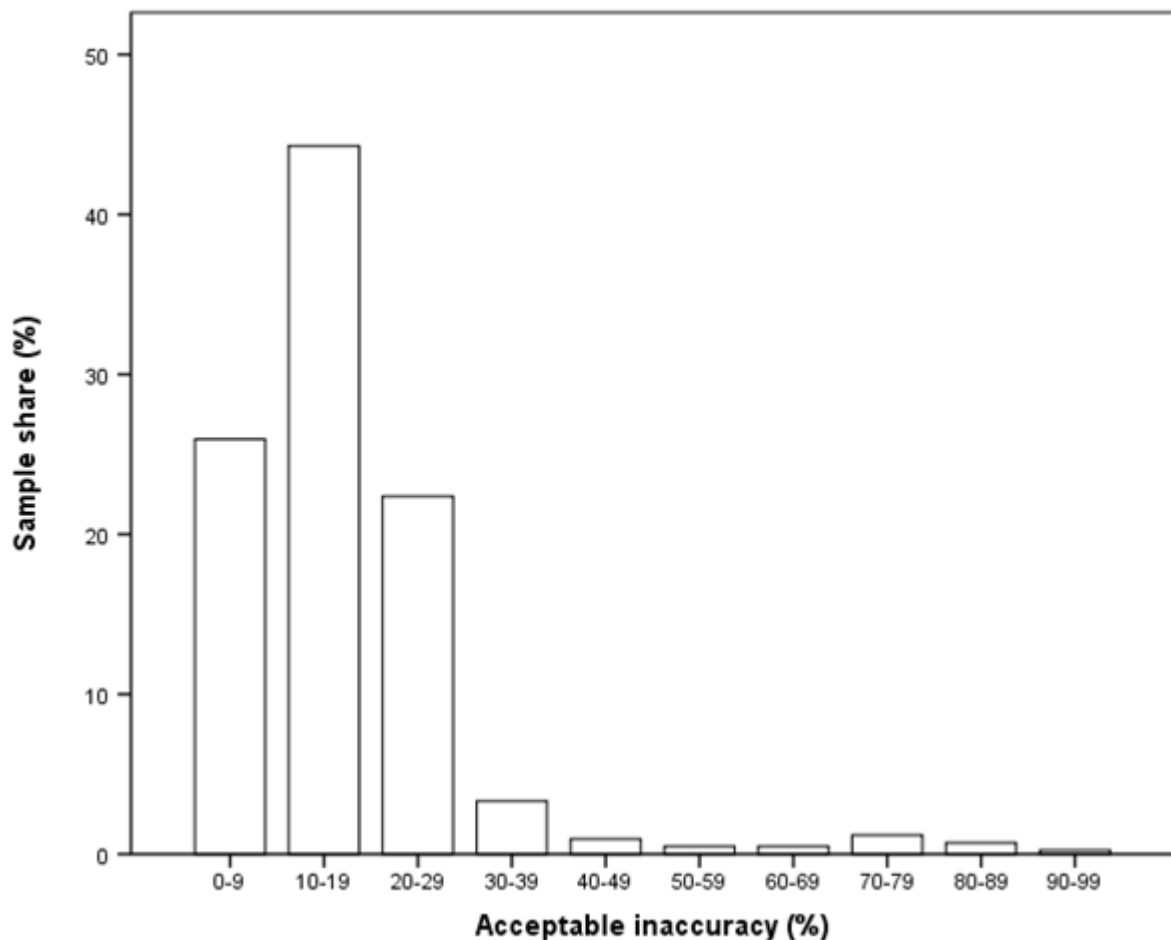


Figure 46: Acceptable inaccuracy in demand forecasts among survey respondents (N=420).

7.2.1 Perception of uncertainty and bias

The results reported in section 7.1 indicate that uncertainty is generally communicated fairly poorly in decision support documents for transport infrastructure projects. As already mentioned, it is possible that uncertainty is handled through more informal procedures, and while this is certainly not optimal it might reduce the impact of certain types of inaccuracy. The present section addresses this issue based on the responses for the questionnaires sent out to stakeholders in Scandinavia. Again, the main focus of the present study has been demand forecasts, and the questionnaire therefore doesn't address issues of cost inaccuracy.

We might start out by looking at how much inaccuracy stakeholders deem reasonable to tolerate. In other words, what is the level of inaccuracy that they find acceptable in travel demand forecasts? We asked this specific question in our questionnaire by allowing respondents to select between different categories of uncertainty spans, and the answers can be seen in Figure 46. We observe that a vast majority of respondents find inaccuracy of outside $\pm 20\%$ span to be unacceptable, and very few respondents tolerate inaccuracies outside the $\pm 30\%$ span. Comparing these levels with the findings from chapter 6 it is clear that in practice considerable uncertainty falls outside the acceptable span indicated by stakeholders. However, it also indicates that most stakeholders acknowledge some of the inherent uncertainty in demand forecasting. There doesn't appear to be any statistically

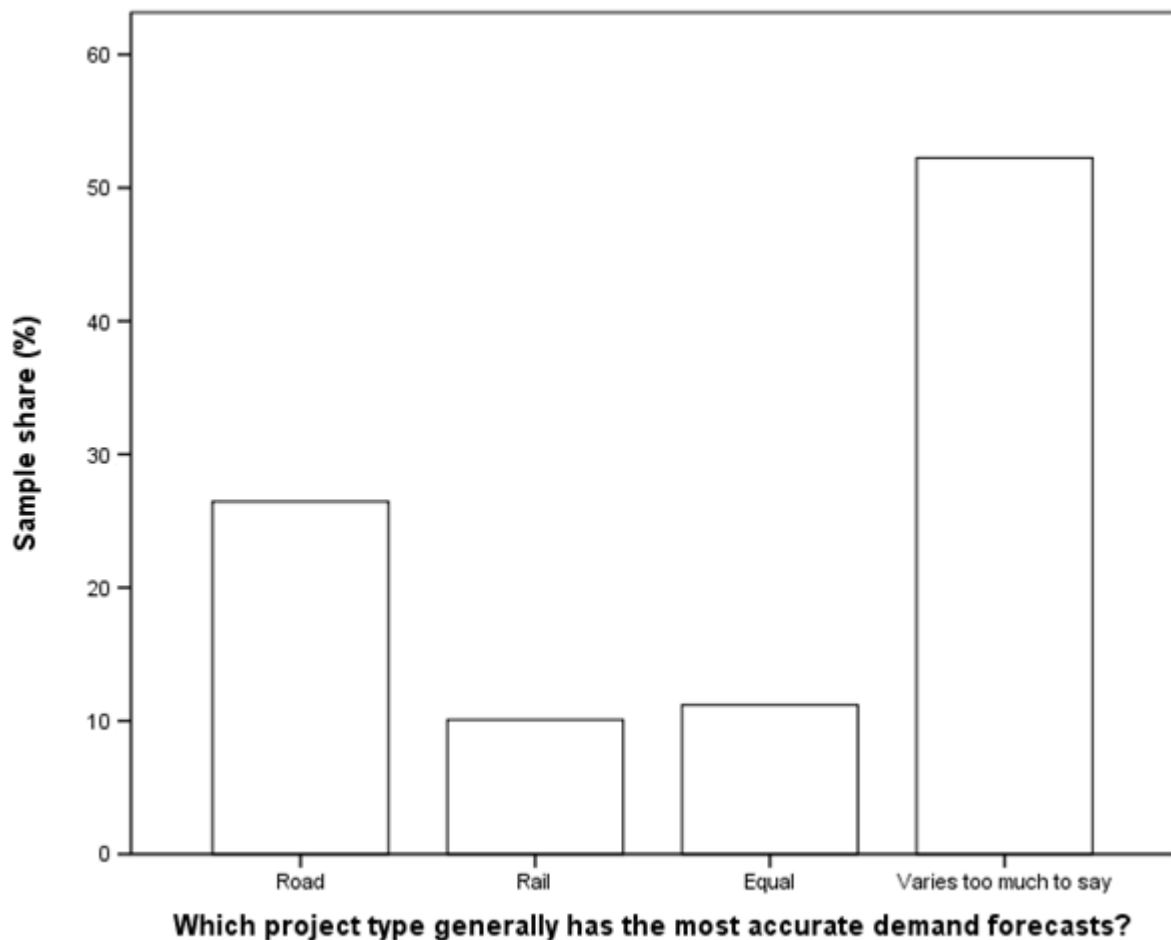


Figure 47: Perceived accuracy of forecasts for different project types among survey respondents (N=446).

significant differences in acceptable levels of uncertainty among various groups of respondents, although there are some weak indications that stakeholders at the local level tolerate higher levels of inaccuracy than those at the national level, just as people who order forecasts appear to tolerate higher levels of inaccuracy than those that produce them.

Next we look at whether stakeholders perceive different levels of accuracy for rail and road projects. We didn't supply any quantified categories for this, but simply asked respondents to indicate which type of projects typically had the most accurate demand forecasts, based on their own experience. The answers can be seen in figure Figure 47 where we observe a set of immediate results. First, an overwhelming majority of respondents indicate that it is not possible to distinguish between the two projects types due to the degree of variation in observed inaccuracy. While not completely in line with the findings from chapter 3 in terms of bias, it at least suggests that many stakeholders experience a considerable degree of inaccuracy in terms of imprecision. Second, among the respondents who do rank road and rail projects in terms of forecasting accuracy, it seems most people consider road projects to have more accurate forecasts. This is in line with the observed inaccuracy in terms of bias from chapter 6. At first glance we might suspect that this could be a result of the majority of respondents working in the road sector, since there are generally many more people working with road based rather than rail based transport. This is true for our sample as well,

Project type	Most accurate project type				Total
	Road	Rail	Equal	Varies too much	
Road	58	13	18	150	239
Rail	6	8	6	23	43
Both	54	24	26	59	163
Total	118	45	50	232	445

Table 13: Crosstab of sector affiliation and perceived accuracy ranking of project types among survey respondents.

but whereas there is no statistically significant difference between the rankings for people working in the rail sector, there is a strong statistically significant difference between the rankings for people working in the road sector as well as people that work in both sectors.

A contingency table with these results based on sector affiliation among respondents can be seen in Table 13. While the results say nothing about the distinction between bias and imprecision, it does suggest that past studies such as those presented in chapter 6 could have raised awareness of forecasting inaccuracy terms of bias for rail projects⁵². This is both comforting and problematic at the same time. Comforting since it means that at least some stakeholders are aware of important research results within the field. Problematic since the large deviations between different studies might cause an undesirably high scepticism of ridership forecasts for rail projects if little consideration is given to the context of the results. For example, comparing the results of forecasting inaccuracy in rail projects from chapters 3 and 6 we could reach very different conclusions about the level of accuracy depending on which study we refer to. This is ultimately a reference class problem, which will be discussed in chapter 8. However, it should not initially be assumed that the average stakeholder has enough background information to make such distinctions between research results.

In the questionnaires we also asked respondents to state their agreement with a series of statements about travel demand forecasts. We used a 5-point Likert scale response format for this (1 = disagree, 5 = agree) to allow for simple numerical tests of the results. In this format we presented the issue of bias more directly, as can be seen from the results in Figure 48. We observe a neat normal distribution around a Likert-score of 3 (neutral). This is somewhat surprising, since I would expect the majority of respondents to disagree with this fairly strong accusation. This distribution is fairly consistent across groups of respondents, and the only group that show a statistically significant difference is people that indicated that they do not work with demand forecasts. The average Likert-score here is surprisingly high compared to people outside this group (3.63 vs. 2.85). This suggests that people who only have limited insight into the production and use of forecasts as decision support are very sceptical of the results, which supports the black-boxing argument put forward by Hajer (1997). Furthermore, among this group of people there is a much stronger agreement on bias for people working with rail projects (3.75) or a combination of road and rail (3.97) than people working only with road projects (3.27). This supports the perception of people generally being more sceptical about demand forecasts for rail than for road projects.

⁵² A few respondents even cited findings from specific studies, including some of those included in chapter 3.

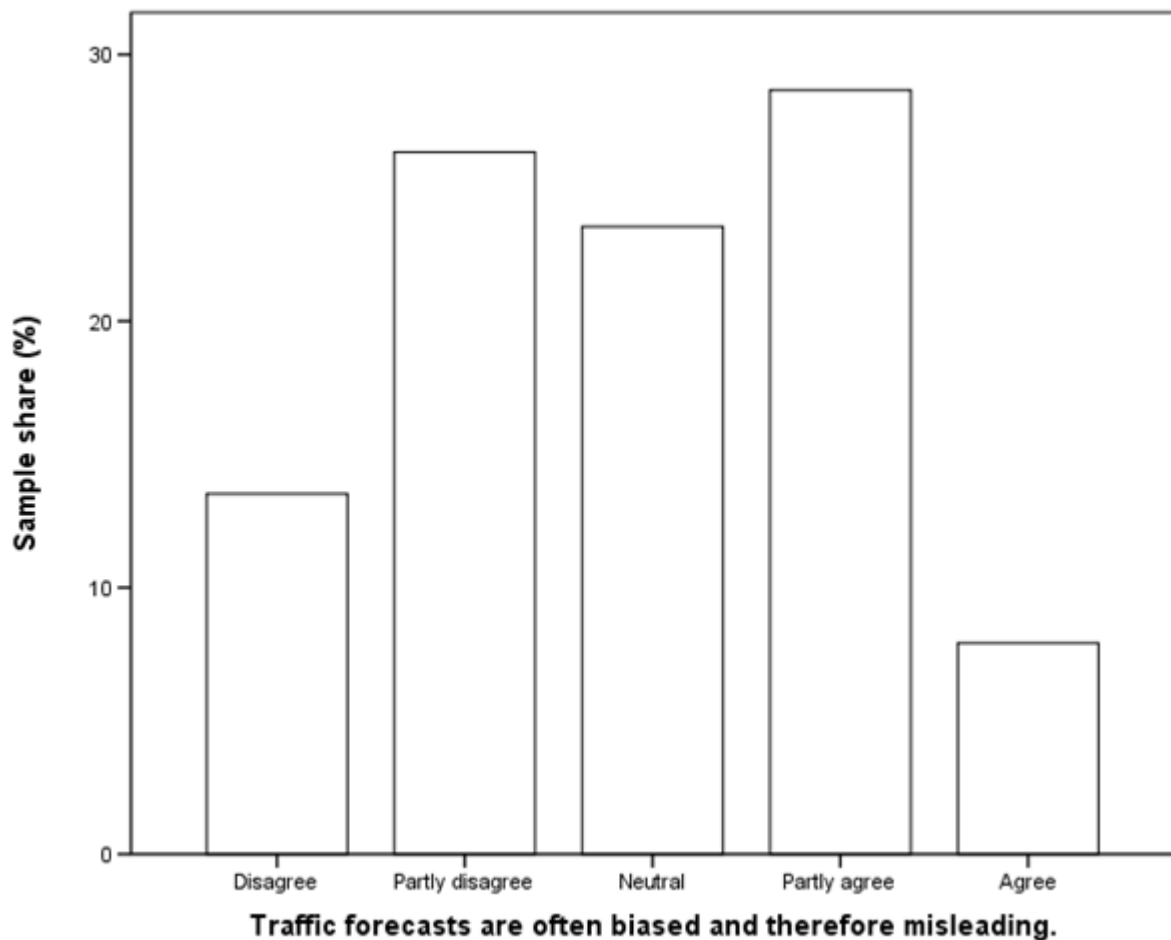


Figure 48: Perception of forecasting bias among survey respondents (N=420).

We also asked the respondents to assess whether forecasting uncertainty was adequately documented in the policy documents where they are used. The results can be seen in Figure 49, where we observe a strong agreement. Considering the findings presented in section 7.1, it would also be quite surprising if stakeholders didn't perceive communication of uncertainty to be an issue. Once again, the distribution is fairly consistent across groups of stakeholders, but two characteristics of respondents seem to stand out. The first is political involvement, since politicians consider the lack of documentation less problematic than non-politicians (3.42 vs. 3.92). Along with the previous observation of scepticism among people who are not part of the production and use of demand forecasts, this gives some support to Mackenzie's (1993) hypothesis of a 'certainty through'. The basic premise of the hypothesis is that knowledge producers are more aware of uncertainty than knowledge users, who tend to "*believe what the brochures tell them*" (Mackenzie 1993, 371). On the other hand, people who are not part of the forecasting regime also consider the knowledge to be more uncertain, often more so than the knowledge producers consider warranted. The second characteristic is that people with higher levels of education is correlated with a more critical perception of uncertainty documentation. People without an academic background have the lowest Likert-score (3.22), followed by bachelors (3.67), masters (3.90) and Ph.D.s (4.34). This is also somewhat unsurprising, given that there is a high covariance between the education level and

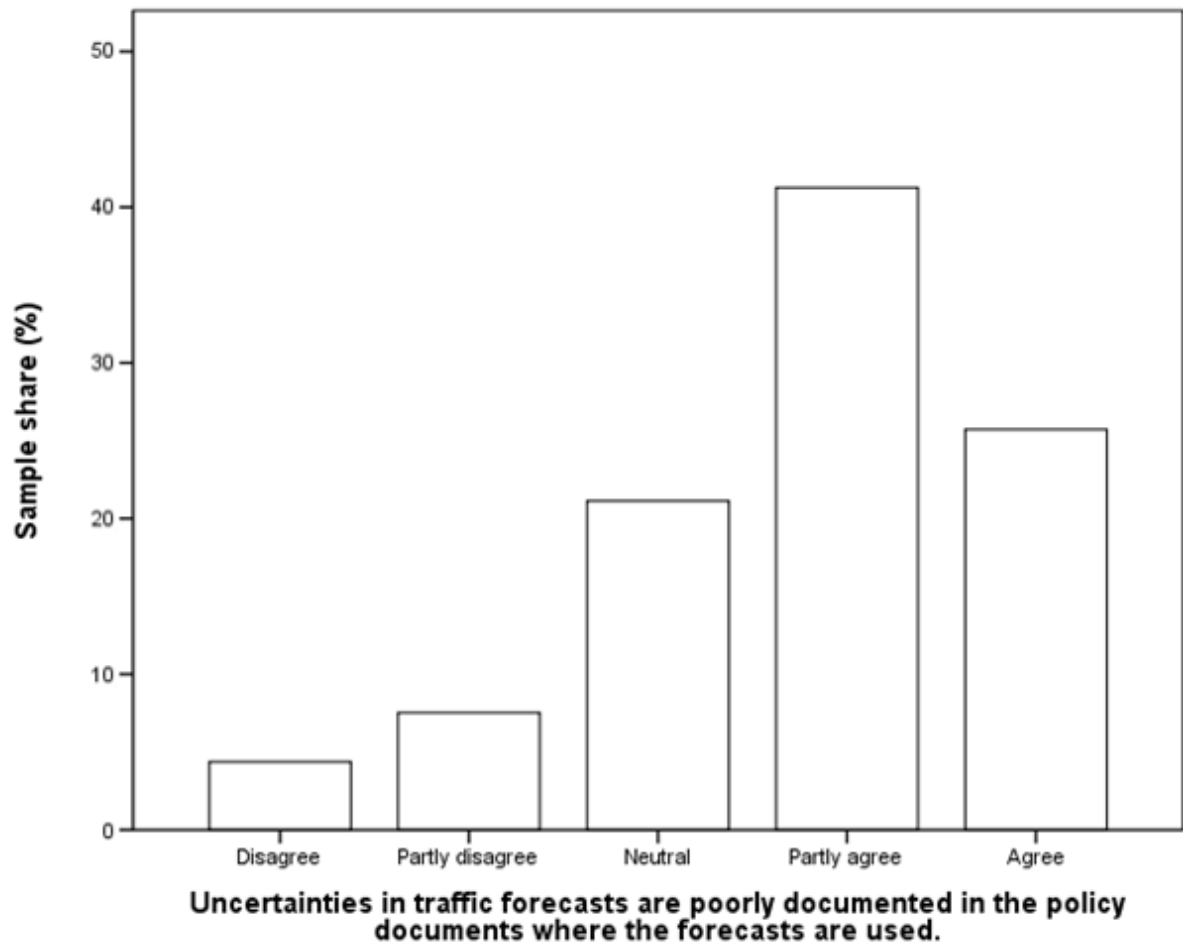


Figure 49: Perception of uncertainty documentation among survey respondents (N=412).

political involvement level. Politicians generally have lower levels of education than non-politicians in our stakeholder survey, since the technical aspects of demand forecasting are typically handled by people with at least a bachelor degree (typically in engineering). However, it highlights the black-boxing problem of model-based decision support where the inherent uncertainty of forecasts appear to be lost in communication between the experts producing them to the non-experts using them, especially if the latter group is not sufficiently aware of this problem.

The black-boxing effect is further emphasized in the results presented in Figure 50. Here we address the respondents' perception of model complexity, and the results clearly indicate that non-experts generally have limited understanding of how the models used in demand forecasting work in practice. When splitting the results up on different groups of respondents we notice that politicians are more confident than non-politicians, which shows through a lower Likert-score on this question (3.25 vs. 3.91). This suggests that politicians might be overconfident in their ability to assess the uncertainty of demand forecasts, since they consider the models to be less complicated than even the people who produce the forecasts. At the other end of the spectrum we have clients (4.55), who are typically government institutions, service operators, or traffic agencies that rely on external consultants for producing demand forecasts. This is interesting, since it essentially means that the

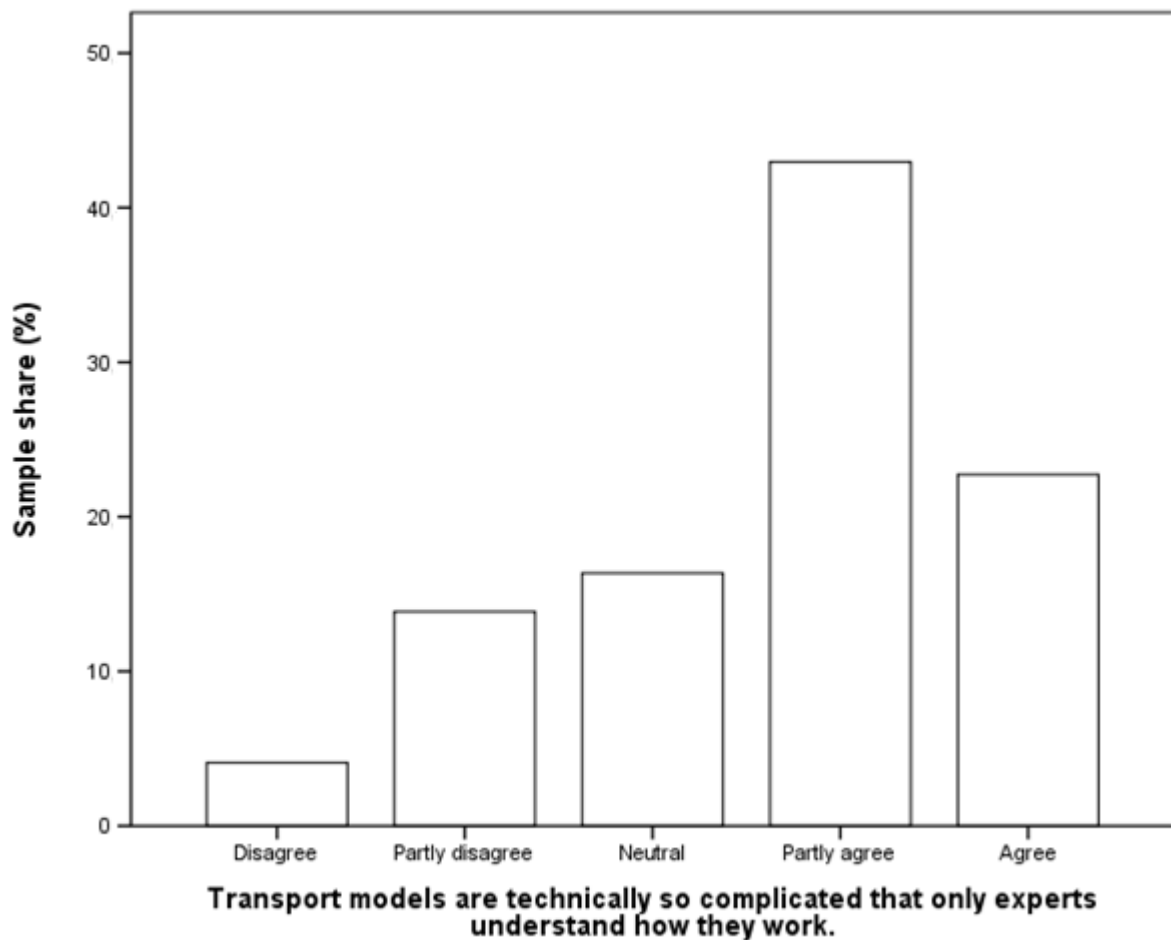


Figure 50: Perception of model complexity among survey respondents (440).

people in charge of negotiating the boundaries of the models are also the ones that consider the black-boxing effect to be the largest.

The above dilemma was illustrated quite nicely during a seminar at Aalborg University in 2011, where a colleague and I brought in a planner from Aalborg Municipality and a traffic consultant from COWI for a guest lecture. The purpose was to discuss the interrelationship between consultants and their clients in matters of forecasting travel demand. The representative from the municipality declared that planners *“blindly trust the numbers we are shown to be correct... we have to, because we don’t really know what goes on behind the curtain”* (Krogh 2011). The consultant agreed, and said quite explicitly that *“modelling takes place at a level where few decision makers are able to keep up”* (Olesen 2011).

As a final note on uncertainty communication we asked respondents whether precise predictions of the future simply weren’t possible. This might seem like a trivial question, and we indeed expected

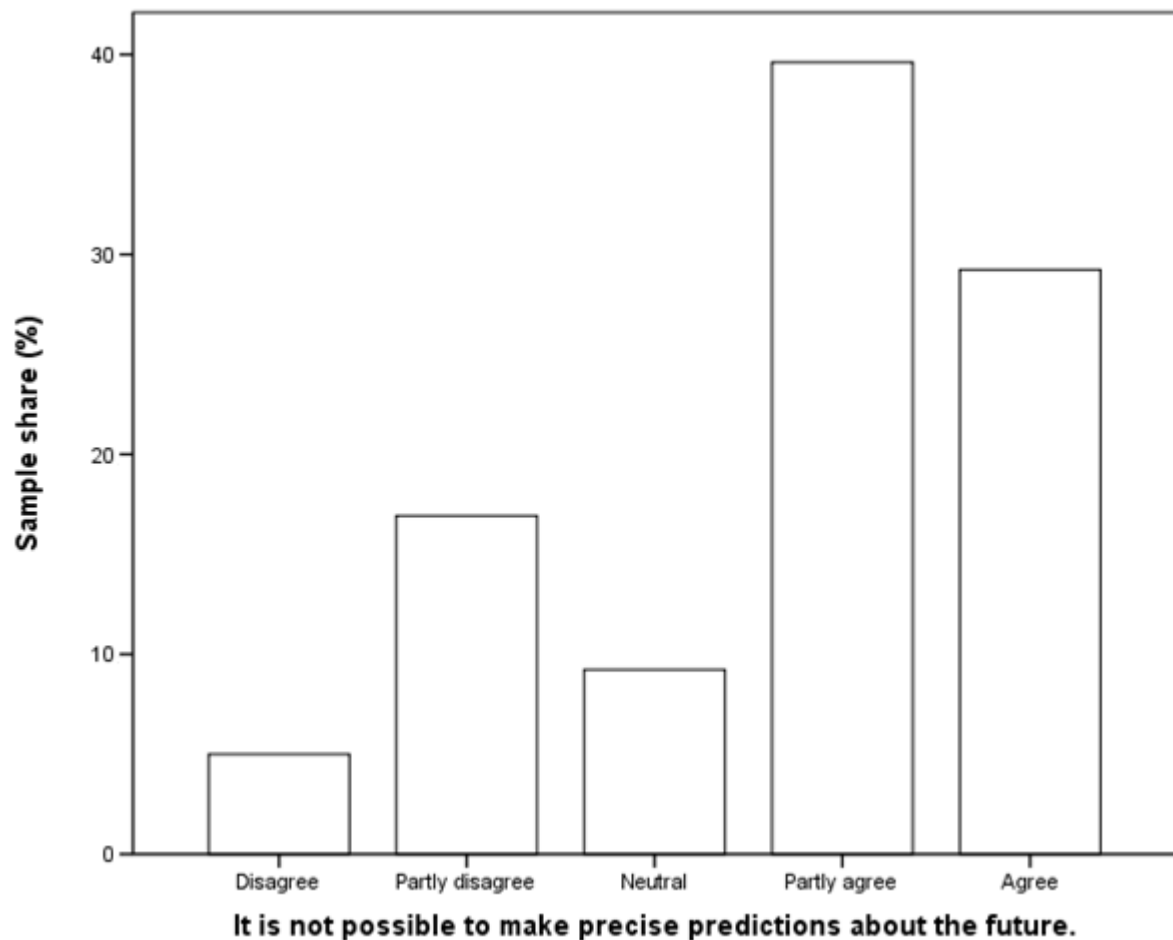


Figure 51: Perception on predictability of the future among survey respondents (N=260).

most people to simply agree with what we considered an obvious statement. As can be seen in Figure 51⁵³, this also turned out to be the case for most respondents. Even so, there are some 20% of respondents who disagrees with this statement, which is somewhat surprising. Of course such a vague statement is open to quite wide interpretation, but claiming that precise predictions of the future is possible sounds a little alarming⁵⁴. It suggests that a relatively large group of people in transport planning adhere to a technocratic planning paradigm, where statistical number crunching is perhaps the most important planning discipline. The results are once again fairly consistent across groups of respondents, but the strategic planning level had some statistically significant differences between different groups. People working at the national level were on average in noticeable less agreement with the statement (3.32) than regional (3.97) and municipal (3.71) planners. Whereas only half the

⁵³ The number of respondents here is somewhat lower since politicians were excluded from this part of the survey. This is an intentional design choice since the question was directed at people who are familiar with the technical capacity of models for demand forecasting. In retrospect it would probably have been useful to include politicians here as well, but the results are still interesting in the interpretation of non-politicians' perception of predictability.

⁵⁴ I refer to chapter 2 for a discussion of why I do not consider this to be possible and why a deterministic perspective is problematic.

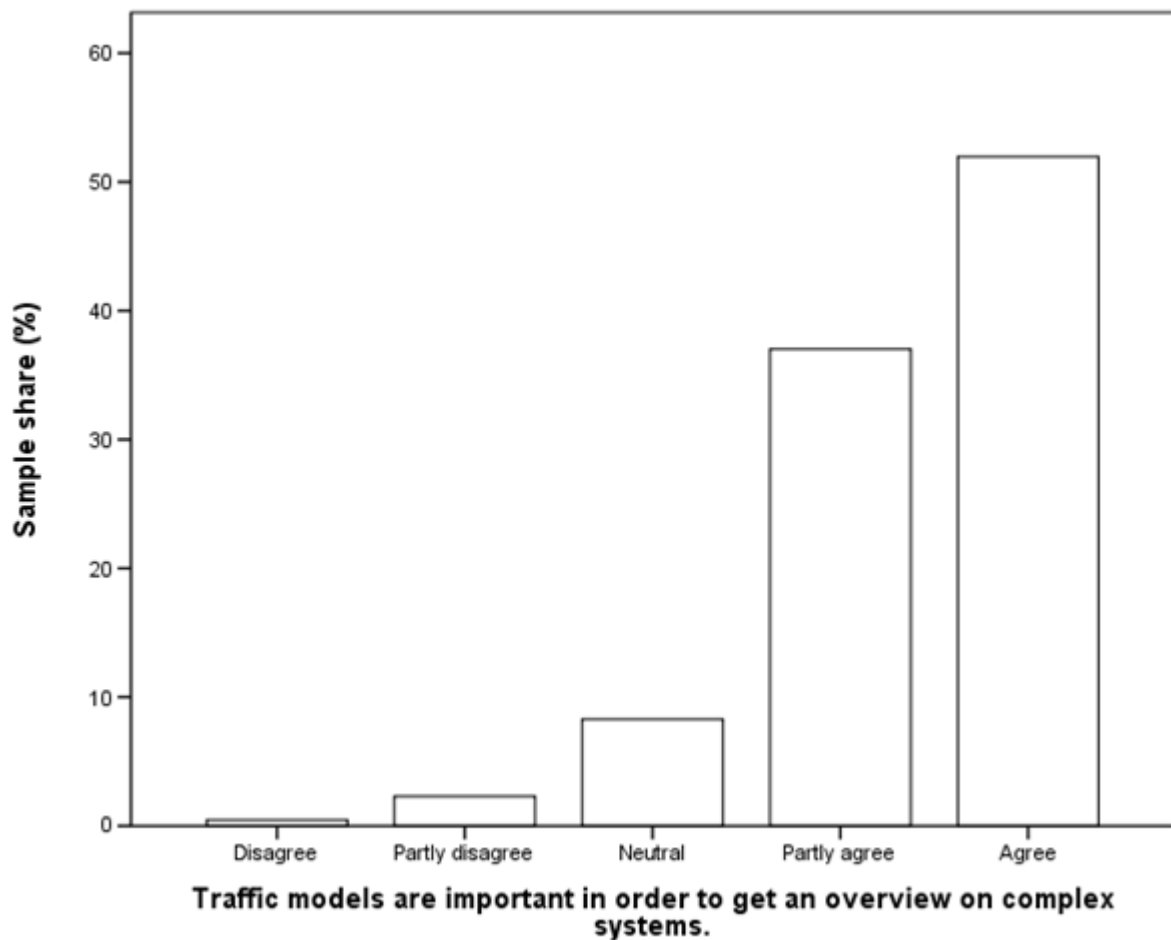


Figure 52: Perception of models as necessary tools for handling system complexity among survey respondents (N=435).

planners at the national level believed precise forecasting to be impossible, this figure was between 70-80% among regional and municipal planners. This is interesting, since one of the primary reasons for forecasting demand is to use the results as input to CBA and EIA to evaluate the value of a project to society. These guidelines are formed at the national planning level, but if the intended purpose of these tools relies on a larger degree of determinism than is actually possible, this weakens the validity of such decision support for transport planning.

7.2.2 The role of demand forecasts

We now turn to look at more general questions directed at the role of models and demand forecasts in policy-making. The purpose of this section is to get an idea of how stakeholders perceive the purpose of the forecasts being produced, and whether the observed inaccuracy and communication of uncertainty is sufficient for these roles.

First we investigate whether stakeholders consider models to be necessary for obtaining an overview over the complex, open system in which transport operates. As discussed in chapter 2, this is typically considered an essential part of the modelling, and we would therefore expect most stakeholders to agree with this. The results can be seen in Figure 52, and it is clear that the vast majority of respondents agree with this statement. The results are consistent across all types of

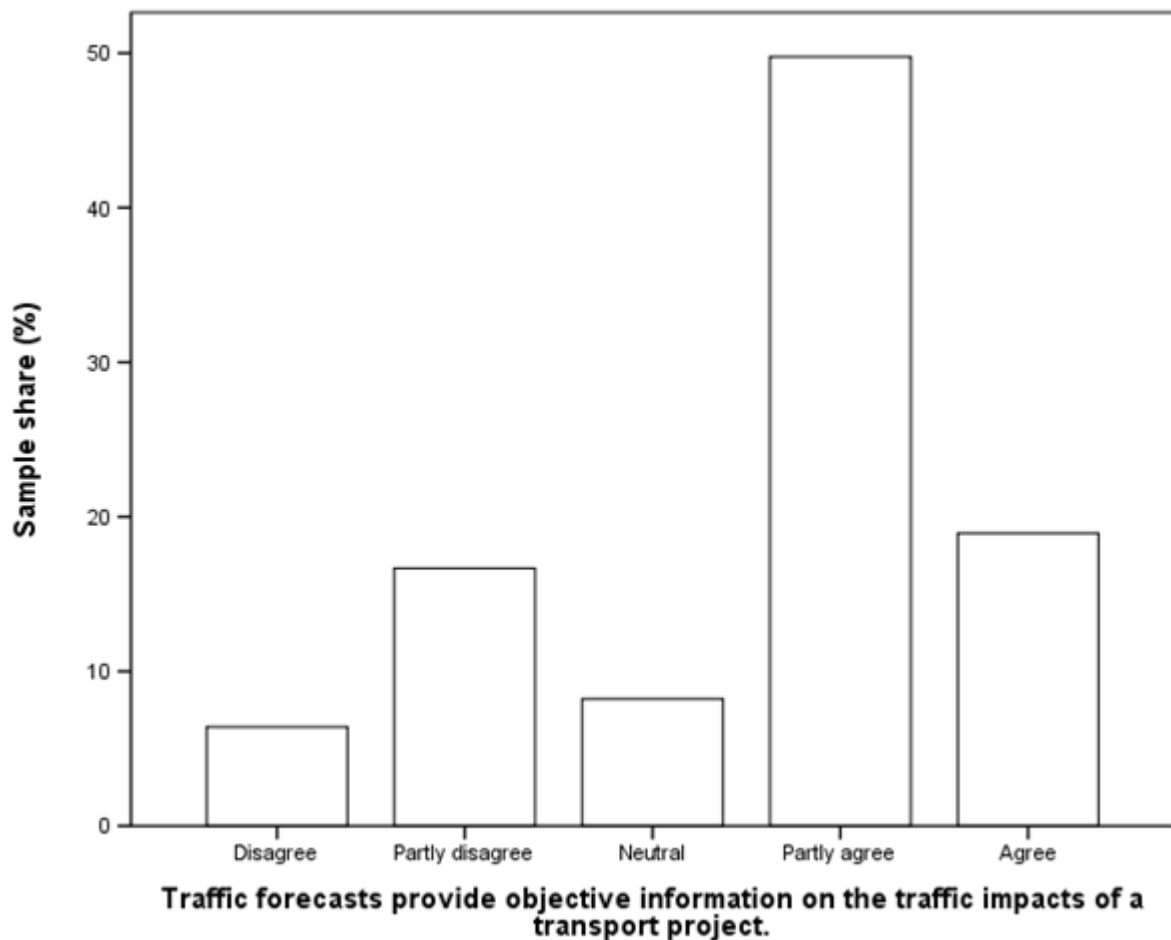


Figure 53: Perceived objectivity of demand forecasts among survey respondents (N=438).

respondents, although there is a statistically significant difference between politicians (4.22) and non-politicians (4.47). However, the difference is quite marginal and certainly does not give rise to any speculation that politicians generally consider transport models to be superfluous to decision-making.

Next we explore how stakeholders perceive the objectivity of demand forecasts. As discussed in chapter 1, the perceived neutrality of cost-benefit analyses is likely an important reason for why it is the dominating appraisal approach, and the purpose is thus partly to assess the extent to which stakeholders are in agreement with this statement. The results can be seen in Figure 53, and once again there is an overwhelming agreement among respondents, although not as extreme as in the previous question. The distribution is consistent across most groups of respondents, but a few statistically significant deviations can be observed. First, consultants agree stronger with this sentiment than non-consultants (3.87 vs. 3.52). This is interesting but perhaps not surprising, since the entire purpose of hiring them is often to receive an objective third-part assessment. Second, people who do not work with demand forecasts agree much less with this sentiment than people who do (2.83 vs. 3.65). This is a remarkably strong deviation given the nature of this question. It supports some of the previous findings that indicate a high scepticism in regard to forecasts among people who are remote to the decision-making process and the employment of forecasts as decision

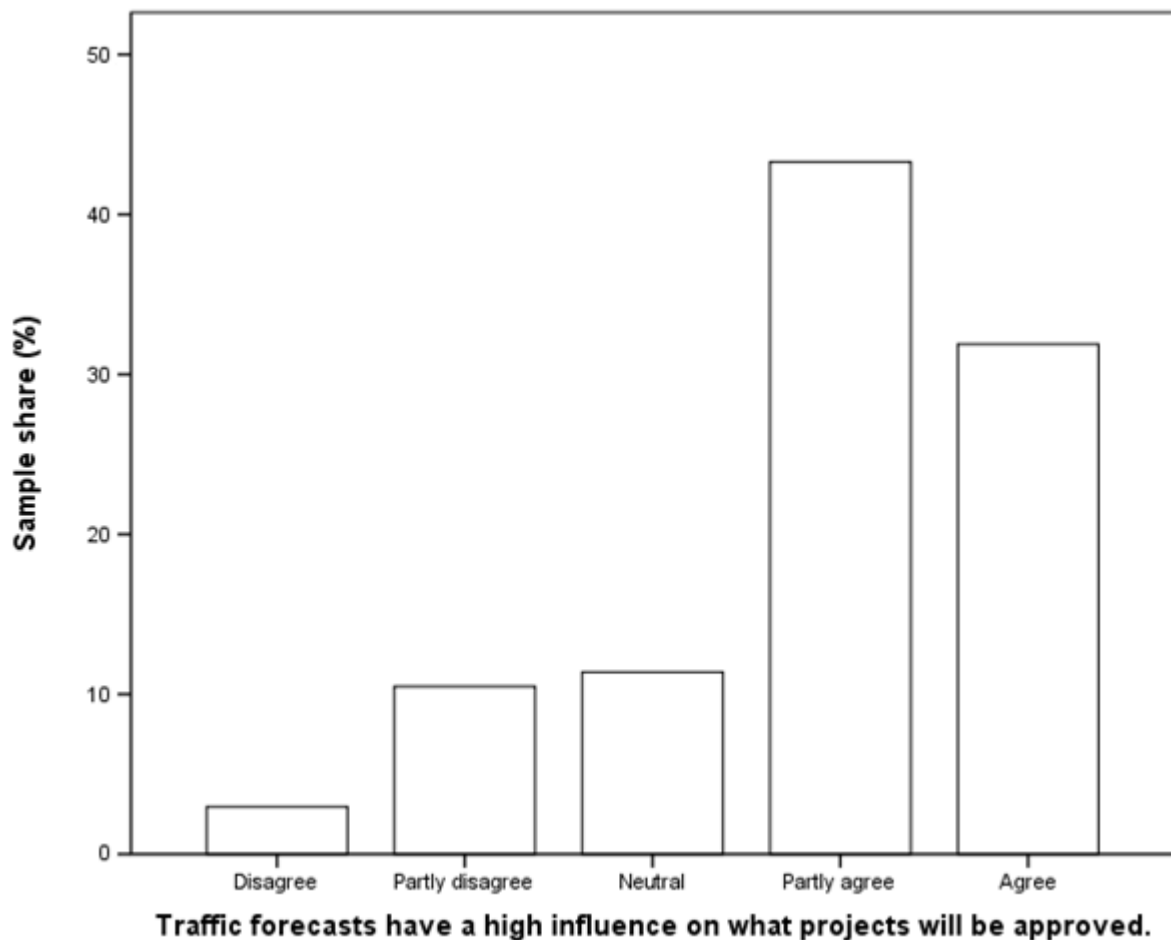


Figure 54: Perceived impact of forecasts on project approval among survey respondents (N=439).

support. Third, there appear to be small deviations based on the respondents' educational backgrounds. On average, a longer education is associated with less perceived objectivity (from 3.78 at bachelor to 3.44 at Ph.D.). In addition, a background in sociology or political science is associated with less perceived objectivity than a background in economics or engineering (3.47 vs. 3.72). This is likely a result of the latter two disciplines being a typical prerequisite for working with the technical aspects of modelling, either in terms of demand forecasts or subsequent impact assessment tools.

So far we have established that most stakeholders consider transport models and demand forecasts to be quite essential decision support tools, but how do they perceive the impact of forecasts on actual decisions? This question was asked quite directly to respondents, and the results can be seen in Figure 54. Once again we observe a high degree of agreement among respondents, indicating that most stakeholders do not only consider the models important to manage complex systems, but the resulting forecasts to exercise a large degree of influence on the outcome of policy-making. In terms of different respondents we find that politicians agree more with this statement than researchers and consultants (4.09 vs. 3.62 and 3.78 respectively). Clients are the group with the highest level of agreement (4.42). Overall it seems that people who work closer to the political arena of decision-making tend to perceive forecasts as having a higher influence on policy-making than people who work with more technical aspects of providing decision support.

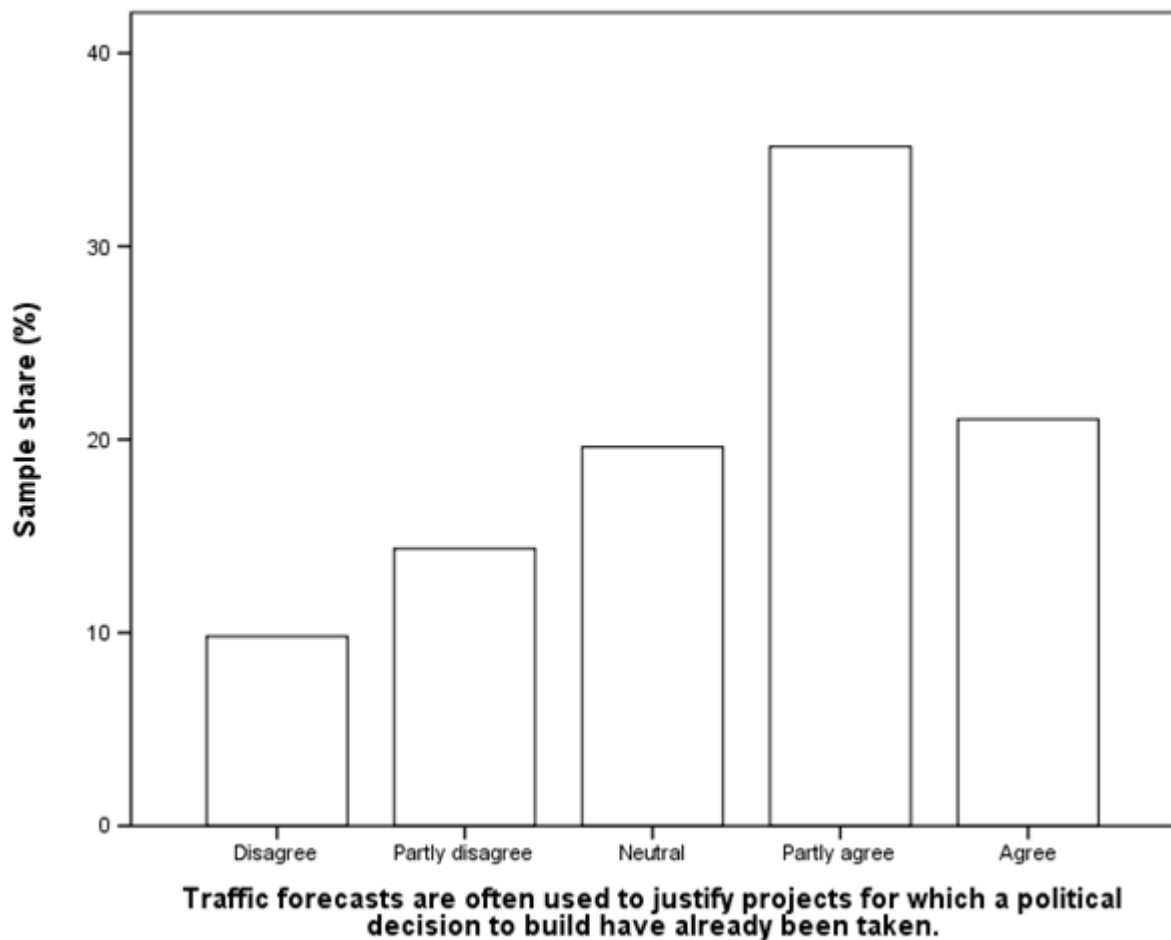


Figure 55: Perception of forecasts used as justification rather than support for projects among survey respondents (N=418).

We now move into more sensitive questions about the more informal roles that forecasts are sometimes accused of taking on. Sceptics often accuse transport models, demand forecasts, and project appraisal to be nothing more than expensive rubber stamps that are necessary to obtain funding⁵⁵. We asked stakeholders if they considered forecasts to be used to justify political decisions that have already been taken in advance, and the results can be seen in Figure 55. We observe that there is a strong agreement to this statement. This can of course be problematic, but perhaps not very surprising. It does not necessarily mean that forecasts cannot be objective or informative tools, but it does suggest a quite strong incentive for achieving a specific outcome that could lead to cognitive bias or intentional deceit. However, there is nothing inherently wrong with justifying an established position via the use of model-based decision support, since this is more or less a necessity under contemporary planning legislation in most countries. The problematic aspect is if there is no learning aspect involved for project proponents when appraising a favoured solution. Rather than using the appraisals as a ranking tool for where to invest, it is likely that most policy-

⁵⁵ As an example, Crompton (2006, 80) cites Shakespeare's *Macbeth* when describing his perspective on economic impact studies, stating that they resemble "*a tale told by an idiot, full of sound and fury, signifying nothing*", but goes on to state that it is perhaps rather the public and politicians who are the idiots, while the tales are told by knowledgeable people who abuse this ignorance.

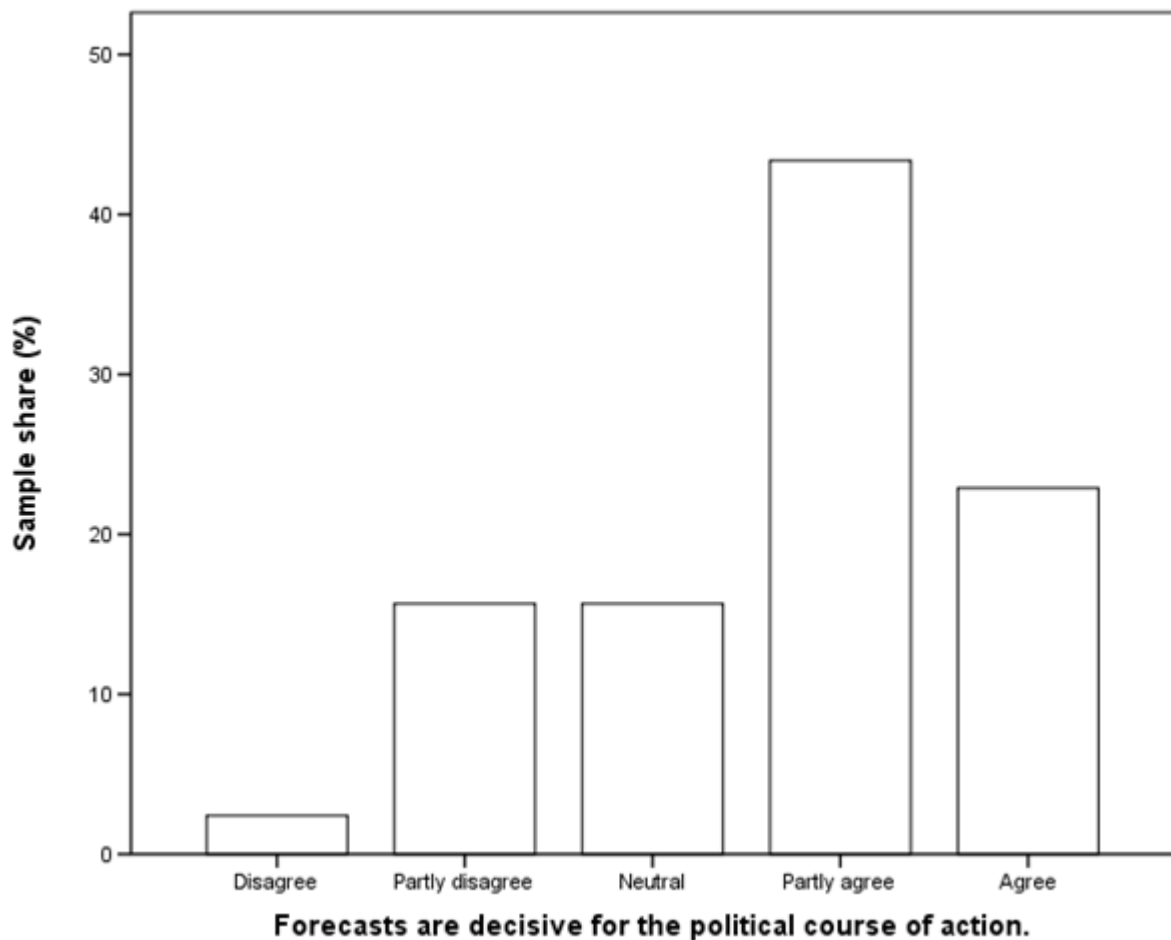


Figure 56: Perception of the decisive role of forecasts in charting the political course of action among survey respondents (N=166).

makers use them as a checks-and-balance measure for projects they already support based on other criteria that might be less tangible. The results are consistent across most groups of respondents, and any significant deviations are too small to be of importance.

The respondents working in the political arena were asked a set of additional questions regarding the use of forecasts in a political context. In the rest of the present section some of the results from these questions are presented, and since we now target a specific subgroup of respondents the number of respondents is notably lower than for the results that have been presented so far. The first question seeks to address whether forecasts are indeed decisive for the political course of action. This is much in line with the question presented in Figure 54, but covering transport planning in general rather than approval of specific projects. The results are presented in Figure 56, and we notice a general trend of agreement with this issue. The only statistically significant difference is among respondents from different strategic levels of planning, with politicians at the national level agreeing much less than those at the regional and municipal level (2.93 vs. 3.82 and 3.68 respectively). The numbers look more drastic than what is likely warranted, since there are relatively

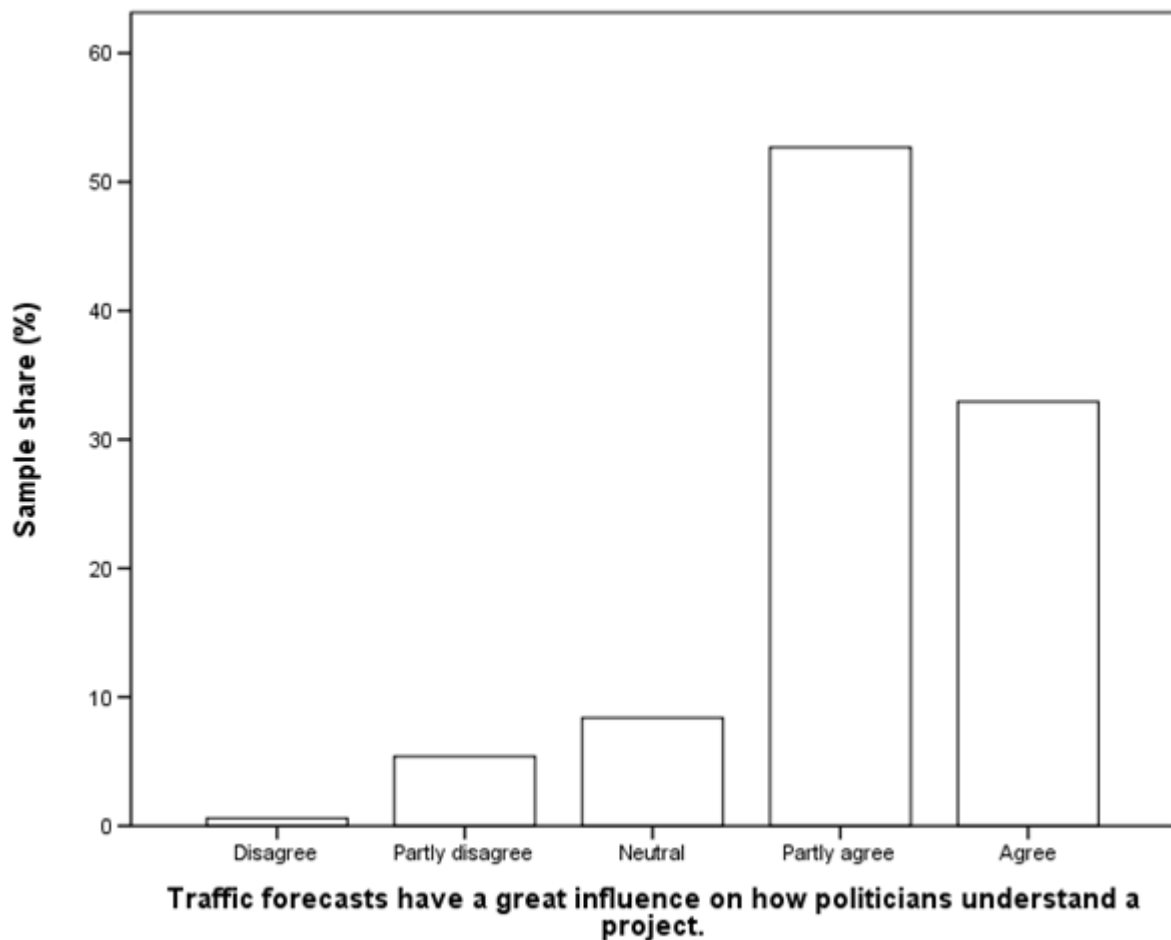


Figure 57: Perception of forecasting influence on the understanding of transport projects (N=166).

few respondents among politicians at the national level compared to those from the regional and municipal levels, resulting in a large standard error (0.33). This is not very surprising given the low population of politicians at the national level working specifically with the transport sector, and we should thus consider this observation a crude tendency rather than compare the average Likert-scores for these groups.

The next question directed specifically at politicians was whether the demand forecasts help them to better understand a project. Figure 57 displays the responses to the statement that this is the case, and we observe a very high level of agreement among stakeholders in forecasts shaping the understanding of projects. In light of the purpose behind modelling discussed in chapter 2, it is quite reassuring that the vast majority of politicians find forecasts to be helpful in this regard. It also means that transport modelling and demand forecasts play an important pedagogic role in framing problems, which needs to be taken into account when the results are employed for decision support. Once again we find a statistically significant difference between respondents at different strategic planning levels, where the national level is associated with less agreement than the regional and municipal level (3.27 vs. 4.17 and 4.25). The standard error is still considerable (0.25) due to the few politicians at the national level in our survey, and the importance of treating the results as tendencies rather than comparing their exact magnitude still applies. Along with the results

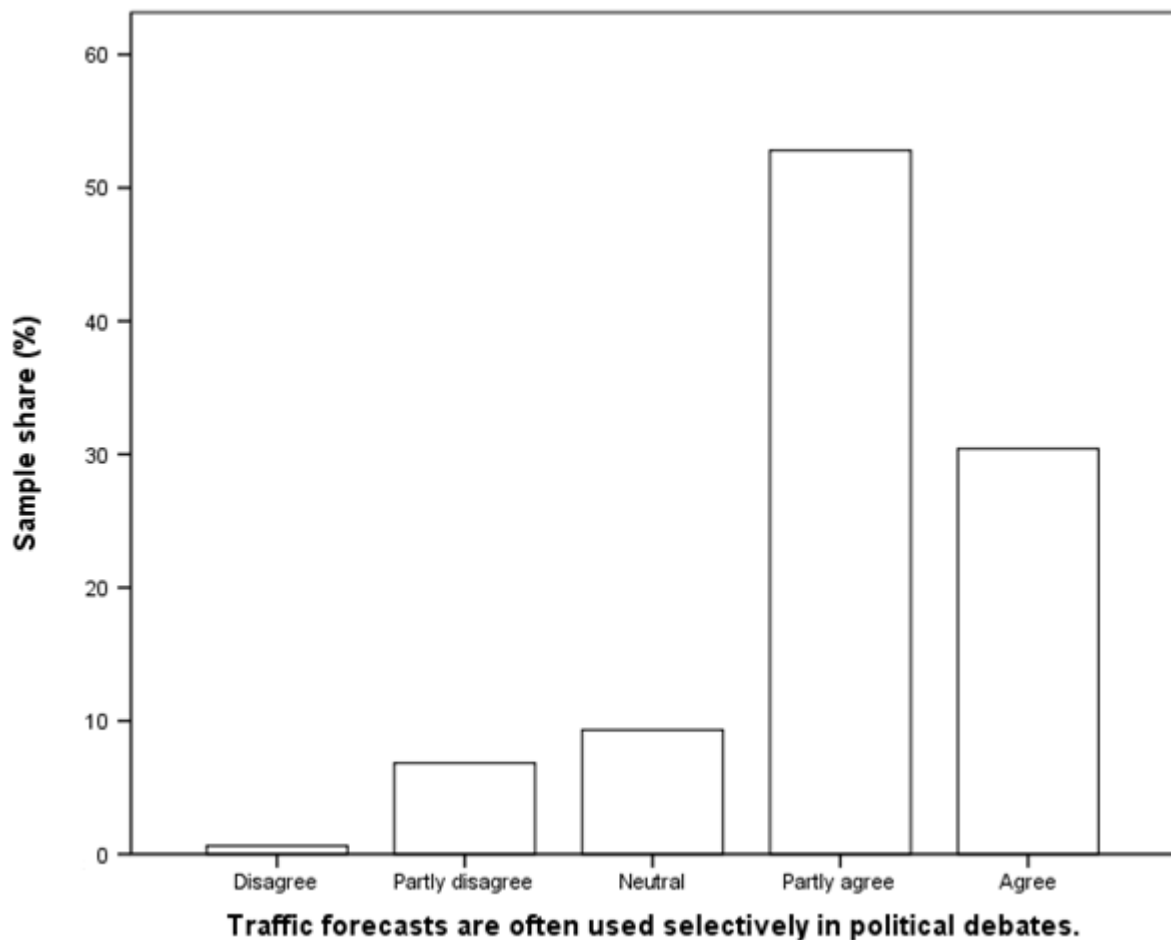


Figure 58: Perception of selective use of forecasts among survey respondents (N=161).

presented in Figure 56 this might be an indication of politicians at the municipal and regional level feeling more dependent on forecasts in order to get local projects accepted due to the vertical bargaining process for state funding or environmental approval. Conversely, politicians at the national level likely face a more horizontal bargaining process that might rely on more informal communication channels.

We also asked the politicians whether they considered forecasts to be used selectively in a political context. The results can be seen in Figure 58, with most respondents agreeing with this statement. These results are also consistent across all groups of respondents. Along with the results presented in Figure 55, it does raise some concern that forecasts are often used as justification rather than support for decisions. It also suggests that the learning aspect of transport modelling and demand forecasts mainly takes place at the personal level for stakeholders, while the formal decision-making process invites for a more symbolic or strategic use of information⁵⁶ to promote partisan politics. This is not necessarily problematic as long as this is a fairly transparent process, but if it is used to mask uncertainty or different scenario outcomes it is concerning. Kain (1990) provided a nice account of

⁵⁶ I refer to March (1994, 225–227) for a discussion of symbolic use of information in decision making processes, and Naustdalslid and Reitan (1994, 53–56) for a discussion of strategic use.

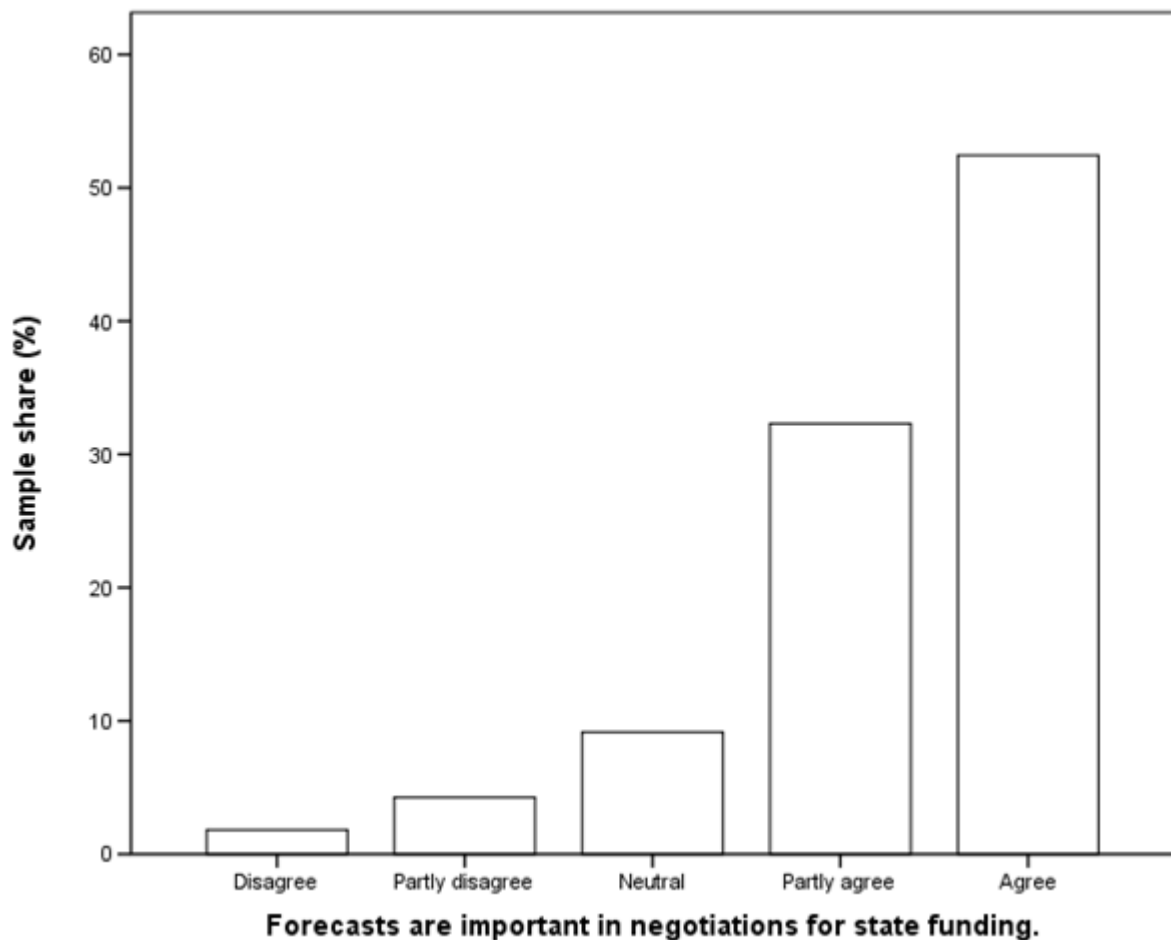


Figure 59: Perception of the role of forecasts in negotiating state funding among survey respondents (N=164).

how selective use of forecasts can be used very deceptively to mislead both politicians and the public, by hiding forecasts of more cost effective solutions that might compete with the preferred alternative by DART/NCTCOG (cf. subsection 3.2.5).

The observations regarding forecasts as decision justification rather than decision support tools could be related to observations regarding the perspectives on the role of forecasts among respondents at different strategic planning levels. We asked politicians whether they considered demand forecasts to be essential in negotiating for state funding, since most large infrastructure projects involve state funding to some degree. The results can be seen in Figure 59. We observe an overwhelming agreement with this being the case, which is probably also to be expected given the dominance of cost-benefit analyses as an appraisal tool. It is common for ministries of finance to have a threshold value for one or more economic performance indicators, and meeting these requirements entails demand forecasts to estimate the majority of benefits included in the cost-benefit appraisals for transport infrastructure projects. In combination with the previous observations put forward in the present section, the results from Figure 59 suggest that project proponents are under pressure to make projects appear as feasible as possible, which might reduce their perceived objectivity.

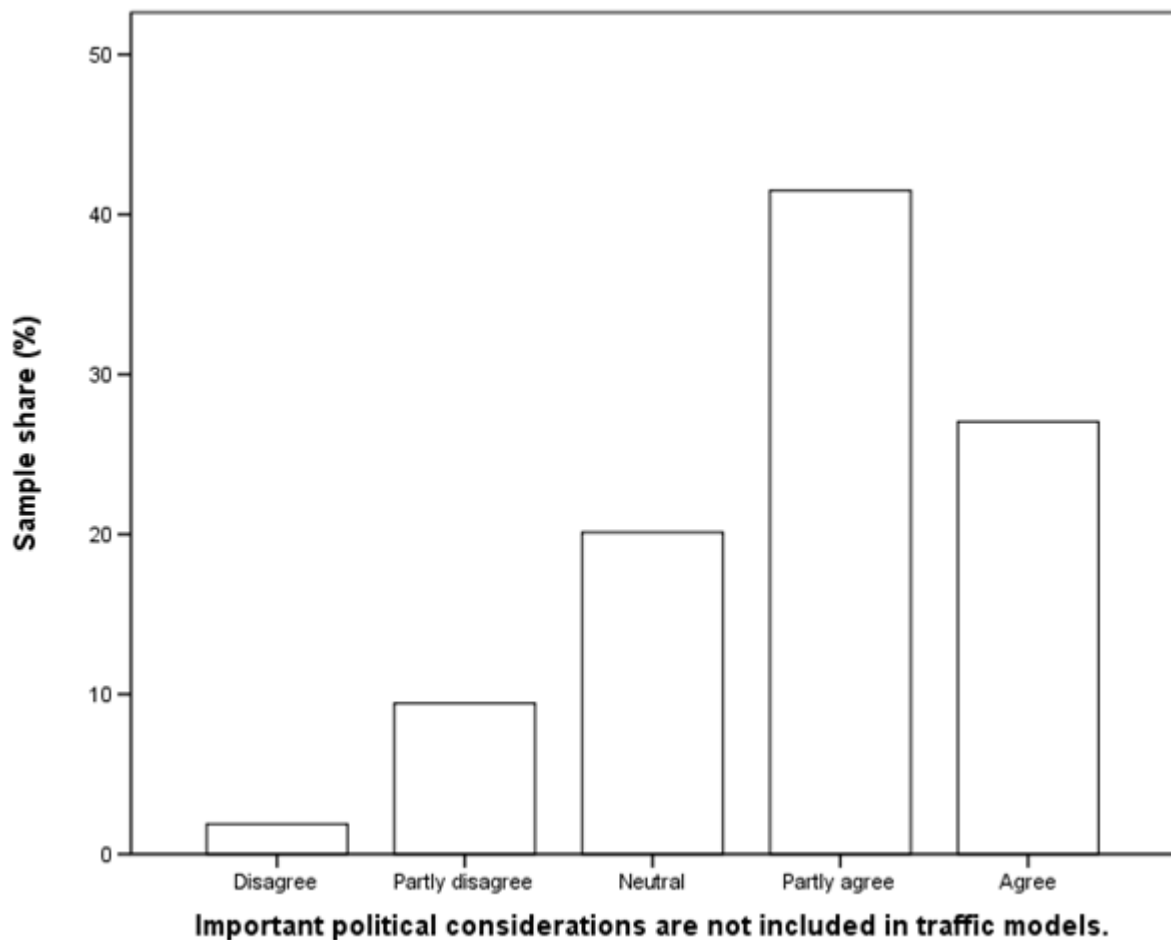


Figure 60: Perception of political priorities being neglected in transport models among survey respondents (N=159).

We also asked politicians whether they considered important political prioritizations to be left out of the modelling tools. The purpose of this question was mainly to gauge how adequate they consider the available models in answering relevant questions that are considered of key importance to policy makers. The results can be seen in Figure 60, and clearly indicate that most respondents consider contemporary models non-exhaustive in the coverage of political priorities related to transport planning. Many stakeholders, both politicians and otherwise, have mentioned this as one of their key concerns with using demand forecasts⁵⁷. They feel that the results presented in traditional analysis give them little or no information on the potential benefits of initiatives to promote demand management or non-motorized transport, just as important impacts such as the distribution of costs and benefits or long-term impacts on land use or neglected. This concern seemed particularly evident among stakeholders who primarily worked with rail projects, and we also find some support for this if we split up respondents based on the project type they primarily work with. Respondents who

⁵⁷ For this claim I mainly refer to my own experiences from discussing the issue with stakeholders, either from personal conversations, email correspondence, in-depth interviews, or statements in the commentary field provided for the questionnaire used in the present study. Much of the evidence is perhaps anecdotal in nature, but I think the observations square well with the experiences of researchers that deal more thoroughly with the topic of stakeholder perception of model-based decision support in transport planning (see e.g. Beukers et al. (2012) for a discussion of problems related to the use of CBA).

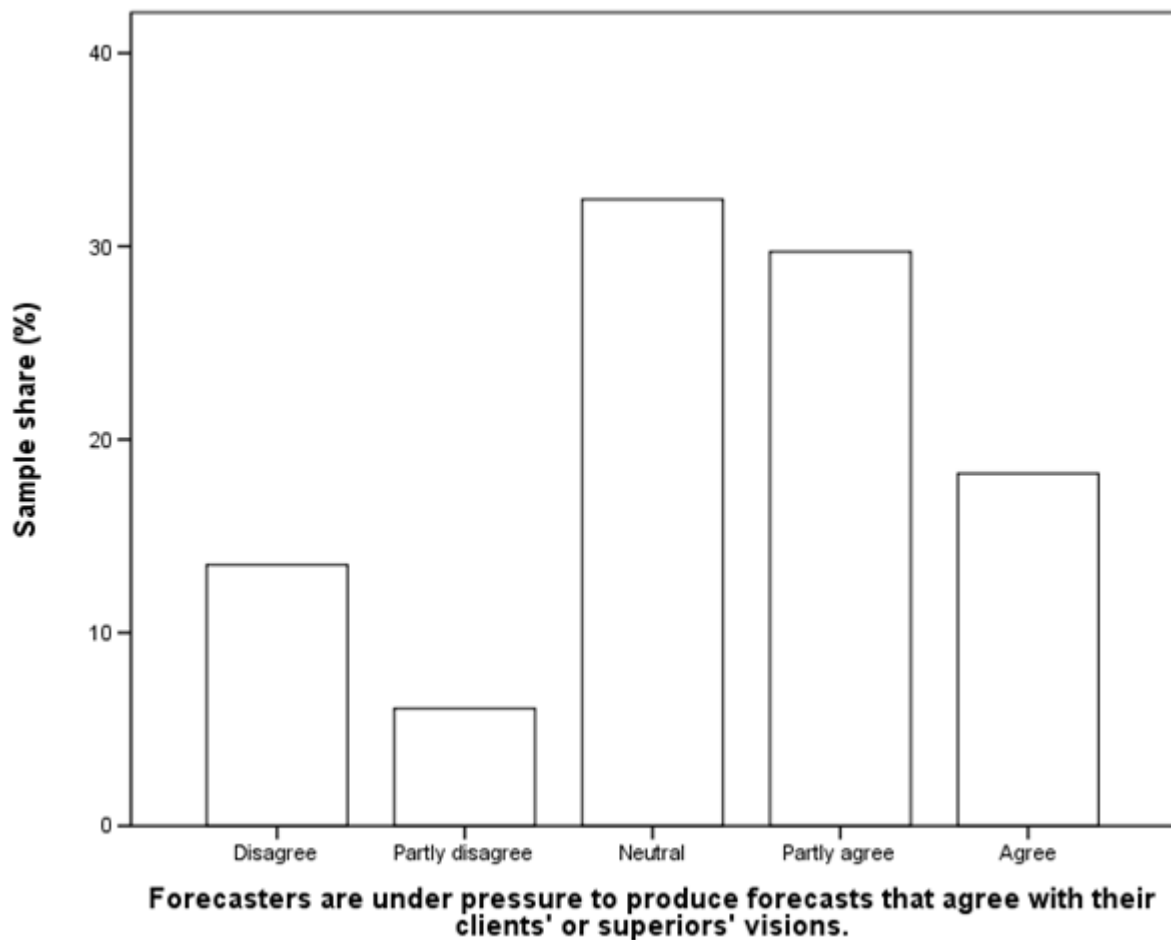


Figure 61: Perception of pressure on forecasters to produce specific results among survey respondents (N=148).

primarily worked with rail projects had the highest average Likert-score (4.43), those who worked with both project was in the middle (4.03), and those working with road project had the lowest score (3.64).

While it is problematic when political debates are obscured by strategic behaviour among politicians, this does not necessarily mean that the demand forecasts themselves should not be considered objective. If a consultant prepares a set of forecasts for a given policy intervention and politicians decide to publically disclose only the forecasts that agree with a specific agenda, this is hardly a fault on part of the planner or a critique of the objectivity of the external consultant. However, if the consultant actively helps promote the politicians agenda, this is very problematic given the black-boxing issues associated with transport modelling. We asked politicians whether they agreed with the statement that forecasters are under pressure to produce results that agree with the visions of their clients. The results can be seen in Figure 61 and indicate a general agreement with this sentiment, although many respondents also indicate to be neutral on the issue. This is somewhat concerning, since the consultants producing the forecasts hold a considerable degree of freedom when it comes to model specification and calibration. Returning to the previously mentioned seminar in which a colleague and I invited a municipal planner and traffic consultant to discuss these issues, the representative from COWI formulated it quite explicitly (Olesen 2011):

The model is a black box – the expert decides the result. They can be used for grave manipulation, and it is the sole responsibility of the expert that the results reflect reality instead of partisan politics.

Of course, being under pressure and succumbing to it are two very different things, but it is clear that consultants have a certain incentive to comply with the wishes of their clients in order to ensure future contracts. This is especially so in cases where the firm is also among the likely candidates for construction contracts if the project is approved. At least in Denmark, the list of consultant firms with the technical expertise to prepare comprehensive decision support is fairly short, and most of them are also involved in construction of transport infrastructure projects.

The observations presented in the present subsection confirms the importance of forecast in communicating needs from local actors through the vertical hierarchy in the decision-making process. Project proponents are thus dependent on transport models to gain an overview over complex social systems, shape their understanding of likely impacts resulting from new transport projects introduced in these systems, and attract financial support to fund these projects. Meanwhile, they do not feel they understand what goes on inside the traffic models, and consider them to be black-boxed number crunching tools that only experts know how to operate; a perception that experts largely agree with. Furthermore, the models fail to reflect important political prioritizations, particularly in the case of rail projects. In this context it is perhaps not surprising that forecasts are used selectively or that stakeholders consider forecasters to be under pressure to produce certain results, since the models appear to have two separate roles depending on the context they are used within. For the individual stakeholder they can act as learning tools, to help gain a better understanding of a complex problem. However, in the decision-making process as a whole the models are instead valued for their perceived objectivity, and thus become an important (and necessary) tool for arguing in favour of project approval and potential state funding. At this point the individual stakeholders is less concerned with the learning aspect and more so with the strength of argumentation. At this stage the models become decision justification tools rather than decision support tools, since the individual stakeholder has already completed his learning process and now needs to secure funding for the project.

7.2.3 Explaining inaccuracy

The last part of section 7.2 is aimed at the most suitable explanations for explaining forecasting inaccuracy from the perspective of stakeholders. The analysis presented in the present subsection will not include the full sample of questionnaire respondents, as was also the case for many of the figures presented in subsection 7.2.3. The reasoning is the same as last, except this time the focus is on the non-politicians rather than the politicians due to politicians having a relatively limited frame of reference for the practical details of demand forecasting in practice. The approach to the analysis regarding explanations of forecasting inaccuracy has also been similar, with a variety of explanations offered individually and a Likert-scale response format to rank the level of agreement from each respondent. However, rather than presenting each of these results individually, I have chosen to simply do a comparison of mean values across all the explanations for which the respondents were asked to rank their level agreement. Before presenting the results of this comparison I will briefly describe the offered explanation and the reason for including it as an option.

Induced demand relates to neglect of induced demand in traffic models. I have already covered why such neglect is problematic in relation to demand forecasting in subsection 6.3.2, including the impact on subsequent impact appraisal.

Poor data quality relates to poor quality of data available to modellers when producing traffic forecasts. This primarily covers the necessary data for model calibration, since substantial data upgrades are resource demanding and thus often not very frequent.

Cognitive bias relates to unrealistic optimism among forecasts regarding the potential success of a project. Since the task of forecasting is usually handled by consultancy firms hired by a client with a specific agenda, most forecasters are well aware that one outcome might be more favourable than another for their client, and might unintentionally bias assumptions in line with their clients' wishes.

Design changes relate to major design changes to a project that takes place after project approval. These changes are obviously not accounted for in the decision support prepared prior to this approval, but might themselves be assessed prior to approval in later amendments to a construction law.

Construction delays relate to delays during construction that postpone the project to the extent that conditions specified in demand forecasts no longer apply to the conditions in the opening year, or that have caused travellers to adopt other modes of transport in the meantime.

Manipulated figures relate to the deliberate manipulation of demand forecasts by project proponents in order to portray their favoured solution in a more flattering way than would otherwise be the case. This explanation is mainly defined by its nature of intentional deceit.

Unexpected scenario relates to different external circumstances than the conditions specified in the forecast scenario. This mainly relates to key input variables used to determine expected levels of future travel demand, such as affluence, car ownership, and population.

Land use changes relate to changes in land use development compared to the specified plans at the time of constructing the forecast. Land use is highly related to travel demand and any deviations from planned development will obviously have great impact on the validity of traffic forecasts.

Other transport projects relate to improved transport options via other infrastructure in the same corridor that the analysed project is planned to serve. It is clear that if two projects are constructed separately but end up serving similar purposes for travellers, the associated traffic forecasts are likely to be far off target if the projects are appraised individually.

Ramp up traffic relates to neglect of potential increases of traffic occurring in the years immediately after the opening year in the traffic forecasts. It might take some time for traffic on a new link to saturate as travellers adjust to the new transport options that have been introduced, and since traffic models often assume equilibrium states this effect would result in overestimated traffic volumes.

Client specifications relate to optimistic assumptions specified by the client that might cause an overly optimistic evaluation of the analysed project. This is somewhat related to the manipulated

figures, but does need to be intentional, since clients might unintentionally specify unrealistically optimistic conditions in favour of a project.

All of these were presented on a single page and with more elaborate descriptions in the actual questionnaire, but for the sake of brevity I have assigned them short, descriptive labels for the comparative analysis. The list of available explanations is not meant to be exhaustive or mutually exclusive, but is based on the most commonly encountered explanations from stakeholders, both in available literature and from personal experience during the present study. The list is mainly intended to compare the relative agreement from stakeholders, and any identified shortcomings in this mode of inquiry could form the basis of a discussion for more elaborate investigations. As I argued in section 3.3, it is immensely difficult to identify exact causes of forecasting inaccuracy, and the comparison is thus only a crude indicator of stakeholder perceptions on this issue. Nonetheless, given the apparent lack of systematic investigation of such matters, I do not see any suitable alternative for a large sample of projects in any state-of-the-art research I have come across, and the results should thus be considered a point of departure for future enquiry.

Figure 62 presents the mean Likert-score for each explanation as well as error bars for a 95% confidence interval of the mean. The results seem fairly consistent across different groups of respondents. The only strong statistically significant deviations are for client specification and unexpected scenario, where the respondents' primary experience with specific project types has an influence. Respondents who work with both road and rail projects consider client specification to be a more plausible explanation than respondents who primarily work with one type of projects. Respondents who work with both road and rail or primarily rail find an unexpected scenario to be a more plausible explanation than respondents who primarily work with road projects. I shall return to this observation later, but overall it seems fair to conclude that the results presented in Figure 62 are fairly representative for the group of respondents as a whole.

From these observations we notice that the explanations split roughly into four main groups:

1. Weak support: Manipulated figures
2. Indifferent: Cognitive bias, construction delays, ramp up traffic, client specifications
3. Some support: Induced demand, poor data quality, design changes, land use changes, other transport projects
4. Strong support: Unexpected scenario

Of immediate interest is of course the fact that only one explanation finds weak support, which also happens to be the most controversial of the offered explanations. Of course, if deliberate manipulation is really a main source of forecasting inaccuracy one could argue that most stakeholders would still deny it simply because they do not wish to admit to such shady activities being widespread in the justification of spending large amounts of taxpayers' money. This interpretation is in line with Ricoeur's (1975) notion of 'hermeneutics of suspicion', which has also been pointed out by Osland and Strand (2010) in relation to the general hypothesis of strategic misrepresentation. According to this interpretation we would find support for deceptive behaviour

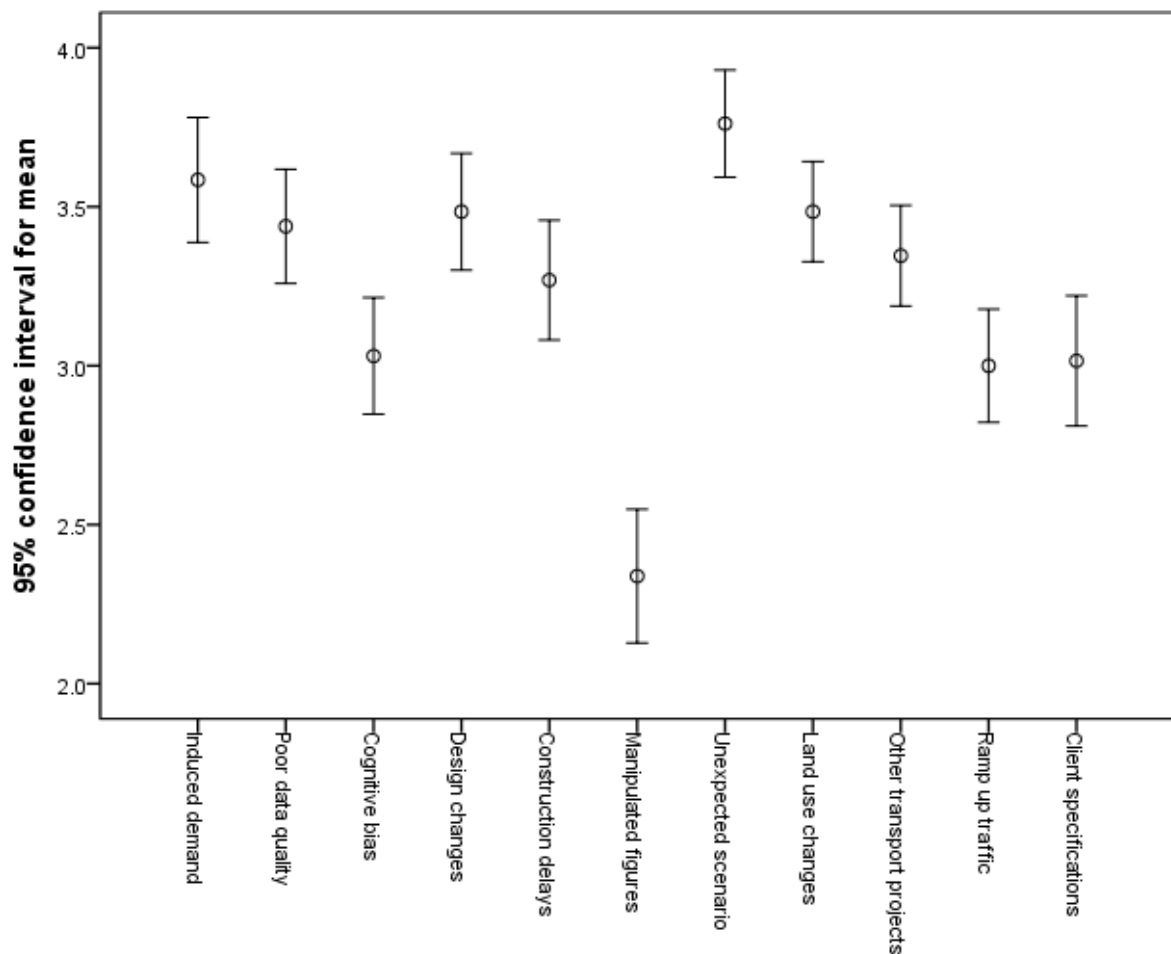


Figure 62: Comparison of mean Likert-scores for explanations of forecasting inaccuracy among survey respondents (N=128).

no matter what level of agreement we might observe from respondents. If respondents agree it is perceived as evidence that the system is rotten, and if they don't it is perceived as evidence of a widespread cover-up. It may indeed be the case that those respondents who agree are only the tip of the iceberg, but it could just as easily be the case that there are simply a few rotten apples in the basket. I don't mean to argue that misrepresentation does not occur at some scale or that it is not a problem of potentially great magnitude, but simply that we should not overanalyse the available evidence to support a theory for which it provides little support. In any case, proponents of the hermeneutics of suspicion must find other ways to support their claims than merely interpreting all evidence as support of their position, as there would otherwise be no way of falsifying their claims. If the responsible stakeholders were indeed trying to cover up their deceptive actions, we would expect to see larger differences among the various groups of stakeholders, which does not appear to be the case. I thus see no indication of a difference between auditors and researchers in terms of favouring specific explanations for forecasting inaccuracy in regard to travel demand, as opposed to Siemiatycki's (2009) findings regarding cost appraisal. Stakeholders generally do not seem to consider strategic behaviour a very dominant explanation.

The group of explanations that on average received more neutral levels of support from stakeholders includes cognitive biases, which might also be a difficult thing to assess in hindsight. After all, this cause of inaccuracy operates unintentionally and likely often unnoticed, and a specific type of cognitive bias is to overestimate one's own ability to grasp complex problems. Like the hermeneutics of suspicion, this type of argumentation risks entering a circular mode of reasoning, and all we can conclude is therefore that stakeholders do not consider this a particularly great risk. Construction delays, ramp up traffic, and client specifications are not considered particularly strong explanations among stakeholders either. The last of these connects a bit with the manipulated figures, as the source of manipulation can take many forms. One of these is through clients specifying favourable conditions for their projects, and here we might recall that respondents working with both road and rail projects had a higher average Likert-score (3.25) vs. those working mainly with a single project type (2.95). This could be an indication that people working with a broader spectrum of projects experience that appraisal for individual project types tend to overlook synergies and conflicts with other project types. We saw an example of this with Gardermoen in section 6.2.5, where client specifications completely overlooked the possibility of competing bus lines for a new high speed rail connection. Construction delays and ramp up traffic are interesting to find in this category, as there has been some claims in academic literature of these explanations being some of the typical explanations offered by planners (see e.g. Bain 2009; Flyvbjerg 2007; Welde and Odeck 2011). While they might often be offered as explanations, stakeholders do not seem to consider them fairly strong in accounting for forecasting inaccuracy.

The group of explanations with some stakeholder support includes induced demand, poor data quality, design changes, land use changes, and other transport projects. Induced demand scores the highest among these, and the standard error for this explanation is much smaller for respondents who only work with road projects than for the entire sample. This indicates a strong agreement with neglect of induced demand being a problem in transport project appraisal for the type of project where it is perhaps most problematic, as discussed in section 6.3. Poor data quality as an explanation squares well with observations from the analysis of decision support documents, in the rare cases that the model specifications are described in detail. Often the data used for model calibration is based on previous studies, since it is both expensive and time consuming to gather new data. It is not uncommon for data used in decision support to be gathered more than a decade ago. Design and land use changes as well as other transport infrastructure are of course crucial to the validity of traffic forecasts, and if these are truly main explanations for forecasting inaccuracy more research should be devoted towards how to deal with these problems. Cantarelli et al. (2012) found that design changes prior to construction are more important for cost overruns than changes after construction has started, and similar arguments could likely be made for demand estimates. This is hardly surprising however, as it is obviously more difficult to make grand changes to project design after construction has started. Rather, it seems that the main problem is a desire to prepare forecasts for a limited set of solutions in a limited set of scenarios, without accepting the fact that both the final solution, as well as the external variables used to determine traffic demand, is part of a span of possible outcomes.

The explanation that finds the strongest amount of support among stakeholders is an unexpected scenario for critical input variables. These variables are (typically) external to the traffic model, and

thus forecasters are rarely asked to nor have the necessary expertise to assess the plausibility of the future values assigned to these values. The findings from several of the previous studies of forecasting accuracy described in chapter 3 also indicate that the development of critical input values often explain large parts of the observed inaccuracies. We might recall that respondents who worked with both road and rail projects as well as those who worked with primarily rail projects had higher average Likert-scores (3.93 and 4.00 respectively) than those primarily working with road projects (3.47). While all of these figures indicate support for unexpected scenarios as a main explanation, the trend is particularly strong among respondents working with rail projects, whether this forms their primary work experience or not. This suggests that rail projects are typically either more sensitive to external variables, more difficult to account for, or based on too optimistic specifications for external variables. It is interesting that many project appraisals are based on only a few or even a single scenario for the development of key input variables, and the findings presented in Figure 62 indicate that a scenario approach to appraisal would probably be more beneficial than attempting to make detailed predictions for the performance under a single scenario.

Overall, the results presented in Figure 62 regarding stakeholder perceptions of the most dominant explanations of forecasting inaccuracy indicate that strategic misrepresentation holds little support among people involved with transport planning in Scandinavia. Rather, they point to unexpected development of external variables, difficulty in assessing behavioural responses, lack of sufficient data quality, design changes after approval, and uncoordinated physical planning as the main causes of forecasting inaccuracy. Still, around 17% of respondents claim that deliberate manipulation is a problem, indicating that the issue is certainly not negligible. It might also be the case that strategic behaviour is disguised in some of the other categories of explanations presented above. Such matters are not trivial to unveil, and should probably be subject to further research. However, even if strategic misrepresentation takes the form of some other category, e.g. optimistic external variables in favour of a given project, the appropriate measure would likely be the same as if this was caused by naïve optimism. More transparent modelling procedures and systematic ex-post evaluation would enable better peer-reviewing of forecasts, reducing both the risk of naïve optimism and strategic behaviour.

7.3 Interviews

In this section I present an analysis of the interviews conducted with stakeholders in Scandinavia. Since many of the survey respondents gave elaborative feedback in the commentary fields, I have decided to use an exploratory analysis of these as a basis for this analysis of the interviews⁵⁸. This is not a departure from the approach outlined in section 4.3, but rather a scoping of concepts based on a source of qualitative data that proved more extensive than I expected in the initial research design. In practice, this was handled with computer assisted qualitative data analysis software (CAQDAS) to obtain a structured overview of the wealth of data that has been obtained through the UNITE project. Emerging issues from the questionnaire comments have been used to form a default set of concepts for coding to reflect the main issues introduced by respondents. These have then been grouped into separate categories to form groups of concepts that relate to each other. Some concepts proved vague or of little relevance to the topic of this thesis, and these have mostly been

⁵⁸ The full comments from the responses to the questionnaire are included in UNITE (2011c).

abandoned. Other concepts proved too general and have been split into more specific sub-concepts. Still others proved redundant, in which cases the concepts were merged. In the end, two dominating categories of concepts emerged (here ranked in order of the number of respondents addressing the concept):

1. Explanation
2. Communication

In the following I will briefly introduce each of these concepts and give an example of a comment addressing each of them. I will not go into detail on the sub-concepts, as these will be elaborated during the analysis of the interviews. The full coding is included in its original form in the UNITE databases for interview and questionnaire responses respectively (2011b, 2011c), which also includes categories and concepts that have been excluded for the present analysis.

The **Explanation** category mainly relates to respondents commenting on how to explain observed forecasting inaccuracies, in extension of the results presented in subsection 7.2.3. An example of a survey response that inspired this category comes from respondent addressing an issue that is often neglected in transport models:

The levels of service, e.g. in the form of time tables for both the planned and existing public transport system, are usually very poorly examined

Another example comes from a respondent addressing a general problem of inadequacy in relation to the models' ability to reflect actual conditions:

The models are poor at integrating effects across different modes of transport. They can't handle heavy traffic very well. Input variables and the assumptions regarding expected development are based on data of very low quality compared to how large an influence they have on results.

Communication mainly relates to respondents commenting on problems in the way knowledge from transport models is produced, processed, and filtered between different stakeholders. An example of a comment labelled under the communication concept comes from a respondent addressing interpretation of model results:

Traffic forecasts often result in the client being surprised about the amount of new traffic a project will generate. This leads to two options: The traditional approach of building new infrastructure or to reduce the demand through mobility management.

Another example comes from a respondent addressing a tendency of stakeholders focusing too much on making accurate predictions, rather than acknowledging the inherent uncertainty of forecasting:

Focus should be on the way we use transport models. In my opinion, it is a waste of time to develop more detailed models – they end up as black boxes or oracles, where no one knows how much of the effect is tied to the large amount of uncertain input variables and how much is tied to actual changes in the transport system. Basic statistical theory

states that a result does not become precise by adding more uncertain parameters. I would rather have a simple but robust model that is able to support the political decision, which a large transport project should always be.

It should be noted that both the overall categories and the associated concepts within them are of course heavily influenced by the framing of the questionnaire, which was specifically aimed at the accuracy of demand forecasts and how stakeholders explain observed inaccuracies. It is therefore no surprise that the explanation category received many comments in the questionnaire, and has the most sub-categories of the four main concepts. Also, many of the comments are labelled in relation to multiple concepts since these are not disjoint classes (nor exhaustive for that matter). Another thing to keep in mind is that stakeholders often refer to forecasts for travel demand and construction costs in a fairly interchangeable manner. This is not necessarily due to a shift in focus during or between interviews, but rather because many informants seemed to consider cost overruns and benefit shortfalls as two related aspects of the same problem. Many informants considered it necessary to decide between additional funding to fix unforeseen problems versus cutting away essential project features to stay on budget. One solution results in cost overruns while the other results in benefit shortfalls.

In the remainder of this section I have split the analysis of interview responses into the two categories identified here. This is mainly for the sake of structure, but obviously there will be some overlap in the relevance of both concepts and citations between the different concepts, just as there is too much interview data to present an exhaustive account of relevant citations. I have tried to avoid reusing citations in these cases and preferred to build upon earlier discussions in cases where an already introduced citation is again of relevance. The two overall categories proved a useful analytic frame for the interviews, but some condensation was necessary in the associated concepts in order to keep the present analysis to a manageable size. The concepts presented under each category are thus themselves categories of additional sub-concepts. A more detailed inquiry of these concepts would be both interesting and fruitful for our purpose here, but could probably form a separate Thesis in and of its own. In addition, the interviews were very openly structured by design. This had the intended effect of informants talking at extensive lengths about the issues they found important in an informal manner, but also to somewhat cluttered arguments and many stops or detours in their descriptions when they lost their train of thought or added new concepts ad hoc. Furthermore, most interviews were conducted in the informant's native tongue to make sure that language was not a barrier in communicating their opinion in detail. I have thus taken the liberty of translating their responses freely when presenting citations, as well as cleaned up some of the most glaring mistakes or confusions in grammar or wording. Most importantly, proper names have often been replaced by place names, as much of the information is sufficiently detailed to allow for identification of the informant by people who have some general knowledge of the issues being discussed. As the interviews were conducted under agreements of confidentiality, I have considered this a necessary precaution, although it is a thin line of judgement that relies solely on my interpretation of the sensitivity of specific pieces of information. I would much prefer a full disclosure of information, since this allows informed readers to conduct critical scrutiny of the statements as well as relate to specific cases being discussed. Nonetheless, much of the information presented here might not have been obtained if now agreement of confidentiality had been established, and I have

thus given priority to the anonymity of respondents. At any rate, all the original material is archived in its raw format to make a peer review of these interpretations possible in cases that require this.

7.3.1 Explanation

As discussed in section 4.3, a disjoint categorisation of explanations for observed inaccuracies in forecasts is somewhat problematic. The responses we got from informants during the interviews reflected this issue well, as many of them offered many different types of explanations and mentioned how one type could easily be the result of another. Even when informants referred to specific projects they had been involved in, they were very reluctant to point to a specific explanation as the main cause of observed inaccuracies. However, they would gladly list what they considered to be likely influences, without making any claims as to the degree of influence on the total inaccuracy.

Neglected effects

One of the most frequently mentioned issues by informants was a neglect of important effects in models, either because these effects were not considered crucial to the overall validity of assessment or because adequate data to model the effects has not been available to forecasters. This is in line with the results for induced demand from section 7.2, but rather than focusing on one particular effect, informants listed several problematic aspects in contemporary modelling. Several informants mentioned a typical neglect of public transport and non-motorized transport; especially in models that operate in areas outside the larger urban areas⁵⁹. One researcher, who had previously worked as a private consultant, explained the issue as a result of path-dependency:

It's a pure car traffic model that is used in this city. I mean, that was what it was designed for. In another city they are building a new model that has both bike and public transport. The old model there is the same as the one here. The models were built by the same companies. But now they have built a bike model and public transport model on top of it. I still work a bit for them, and the first results I have seen for public transport are basically useless. It has to do with the fact that it is not included in the original model but rather built on top. You are able to produce results, but they don't reflect reality. They are configuring and calibrating the model of course, so they'll probably arrive at something useful at some point.

The point being made here is that there is a considerable path dependency in transport modelling, since models are often resource demanding to construct and maintain. The result is that a model is used extensively over a long period once it is built, and rather than build a new model from scratch when new features are required, the existing model is simply adjusted. However, to the informant this is not an optimal solution, and results in a model that is difficult to calibrate. This perspective is supported by a planner in a public transport company in another city, who also claimed that modelling of public transport options was a later addition to the existing car based model for the area:

⁵⁹ This concept is never specified explicitly, but informants' case references indicate that larger urban areas typically refer to cities or metropolitan areas of at least half a million inhabitants and large variety in available transport modes. Another typical indicator is the availability of rail-based public transport systems.

We haven't had a combined model that could do both public transport and cars [...] We had two different models – two different systems. One company had the model for road traffic and another company had the other [for public transport]. It was then harmonised, so they could run together. There are some modal split calculations in it. But I would say that we aren't really doing well on the modelling side.

Obviously this path-dependency puts some restrictions on what it is possible to add to an existing model and what level of detail it is possible to embed in such additions. However, more important is perhaps the likelihood of these necessary compromises resulting in a model that is worse at modelling public transport than car-based transport, since this function is not the primary design objective of the model. This means that for a given geographical area, the quality of transport modelling is likely better for car-based transport than for public transport, *ceteris paribus*. Another implication is that forecasters are rarely updated with new developments in transport modelling, since they operate an old model that is simply patched up to meet new planning challenges as they arise.

Another effect that informants claimed to often be neglected was induced traffic, which was also covered in the survey results presented in section 7.2. One planner in a regional district, who had previously worked as a private consultant, explained the available model's shortcoming in this area, how stakeholders were very aware of it, and what implications it had for the forecasts they had to use:

This model that we have is a stationary model. And that's one of the things we could make explicit; the OD-matrix⁶⁰ is not affected by accessibility. That's one of the things you could criticise about the model we use [...] If you make things more accessible, then that generates more traffic. It has always been like that. It also goes the other way; if you get more congestion, then that generates less traffic due to more resistance. And that is not reflected. This is also why we suddenly estimate 37,000 cars in a street that doesn't even have that capacity.

All informants expressed awareness of the concept of induced demand, although the definition of what the concept covers varied somewhat between informants. As the above citation shows, having no capacity restrictions in the model can lead to forecasts that have little or no connection with reality. As long as these shortcomings in models are made explicit to policy makers, this need not necessarily be a problem. However, uncertainties are often not communicated very well in decision support documents (cf. section 7.1), so this would need to happen through other forms of communication. As one of the two main concept categories deals specifically with communication, I shall return to this issue in the relation to the communication category late in the present section.

Most informants pointed out that induced demand is generally ignored in local models, while most regional and national models have been upgraded to account for increased demand resulting from

⁶⁰ OD: Origin-Destination. A static matrix means that changes in the overall number or average length of trips as a result of changes to the system are not modelled. In other words, the model does not account for induced demand.

increased accessibility. A planner in the national road directorate explains the motivation behind the model upgrades:

The last 10-15 years we have had a model that calculates changes in demand [in the capital area], not just route choice. But the rest of the country hasn't really followed suit, because the national model that was developed at the end of the 1990s isn't used since it isn't any good. It has been decided that a new national traffic model will be built, but it will take some years before it is ready... so the last couple of years we have implemented induced demand in some of our regional models, so we are able to account for it [...] Maybe it's because the projects have become larger. You have to admit that new traffic is in fact generated. And you probably can't ignore that a socioeconomic assessment underestimates the benefits if you do not account for induced demand.

A parliamentary politician made a similar comment regarding the effect on impact assessments:

Many politicians have heard about the effect of induced traffic at least. You can get in the slightly paradoxical situation that when the economic arguments for investment are shaky, then we are aware that there'll be induced traffic. Then we are very aware of it. We had a bit of fun on a bridge project, where we knew there would be induced traffic – otherwise it wouldn't be a feasible project. On the other hand, we weren't much in favour of it when the environmental assessment had to be made. It was a bit back and forth.

These arguments run somewhat counter to the effect illustrated with the case of Nordhavnsvej presented in subsection 6.3.2, where additional traffic led to substantially fewer benefits than when this effect was ignored. The reason is that none of these citations refer to projects built in highly congested areas, and their contexts are thus situations in which the additional traffic would not reduce flow on the network (at least not initially). In these cases the additional traffic will indeed lead to a higher benefit assessment, since more travellers benefit from the improved accessibility. However, the Nordhavnsvej case illustrates why this is dangerous rule of thumb to apply, since this relationship is inversed for congested networks in urban areas. Furthermore, the detrimental effect on flow reduces benefits exponentially, meaning that it is quite difficult to assess the actual benefits when induced demand effects are not sufficiently accounted for. A 5% increase in traffic on an uncongested network roughly translates into 2.5% higher benefit (by the rule of half). Conversely, a 5% increase in traffic on a congested network can result in 40% less benefit from travel time savings (cf. subsection 6.3.2).

There could be many reasons why certain effects are neglected in transport models, but the sizeable cost of building and upgrading them seems to be one of the more important. One private consultant argues that clients typically don't have the resources to specify a very detailed model:

You would never attempt to build a state-of-the-art model with the budgets available anywhere outside the capital region. The model we use there probably cost 3-4 million Euros to develop. We don't have those kinds of clients. That's a one off kind of thing. You did it for a large infrastructure project back in the 1990s, where you needed a better tool

to make forecasts [...] You need to remember that as a consultant you can only council. A client might approach us saying: "We have 40,000 Euros, what can we get for that?" Then that is the assignment we design around. We feel that's appropriate. You can do planning on many different levels, right? But we could also say: "Well, it will be pretty uncertain." And if our people think it's necessary to include induced demand effects, then they ought to state this to the client. But then we also have to say: "That's a larger assignment."

A researcher, who has previously worked as a consultant, supports the statement:

We used the model in the 1970s to reach a conclusion regarding alignment. Then we thought: "Well, that's it". That was actually the purposed of this model. But then this discussion has surfaced four times since... and every time you have considered how much you should adjust the model. I don't know how many adjustments have been made. I know that the general problem has been that the civil servants and consultants have wanted to upgrade a lot of stuff. And you've never been allowed to adjust all the things you wanted, because it was simply much too expensive.

It should come as no surprise that the quality of assessment goes hand in hand with the resources devoted to preparing it, including the resources available to construct the necessary tools. However, since the choice of whether to upgrade the available toolbox resides with politicians, the issue of model sophistication is thus often a political responsibility. This is not necessarily a problem for the individual project as long as the uncertainties of using an outdated model are made explicit to stakeholders. However, it becomes somewhat problematic when comparisons are made across projects from different geographical areas that might have different toolboxes available. A modeller describes this problem explicitly:

Over here they built their model in VISUM, which is a very dynamic way of doing it. Another place they built the model from the ground up themselves. In addition to models being different based on location, they are also built at different points in time, for different purposes, etc. It's problematic that we use so many different consultants with different tools. Different software packages allow different possibilities. Different consultants will give you different interpretations and results.

The point being made here is that the available tools vary greatly between different areas, which make it difficult to compare results of impact appraisals between them. The supposed objectivity of impact assessments require the methodology to be consistent, but if different amounts of resources are devoted to transport modelling in different areas this is obviously not the case. The above citation reflects a plea for consistency and reliability, but this might be because this particular informant worked with prioritizing projects at the national level. Informants working with local projects expressed frustration that the rules were too rigid to account for important effects that applied to specific projects.

Another problem in relation to the neglect of induced demand comes from the comparison with the zero-alternative, where the general traffic growth is overestimated when the deterrent effects of

congestion are not taken into account. This was also pointed out in an earlier citation. A planner from the national road directorate explains how this creates self-fulfilling prophecies in forecasts of travel demand for an expansion of a motorway bridge:

I don't know how decisive [induced demand] is. I mean, we have a motorway today and there will still be a motorway. Actually, when we make the projection we have already included the induced demand. Otherwise there would be a negative effect, right? I mean, if we don't expand it, then there is no room for traffic to grow. And then it cannot possibly reach the figures that the [national] forecasts we use to estimate growth indicate. But if we expand it, then we get the growth we expect.

This further complicates assessments of transport infrastructure projects, since the forecasts for the zero-alternative assumes that the project is going to get built, while it is supposed to reflect a situation in which no project is built. The main problem with this approach is that it overestimates new traffic for the zero-alternatives (cf. subsection 6.1.3), which then leads to overly negative assessments of future congestion levels in the case where no new capacity is added to the system. A parliamentary politician comments on how this illogical estimation procedure causes distrust of forecasts, when discussing the political debate surrounding an assessment of an additional bridge project to solve expected congestion on an existing link:

We had the funny situation where traffic on the existing bridge started dropping. That was a bit unfortunate, since [the proponents] had predicted that it would grow and grow; and then it fell instead [...] But their point of departure is that we need to make room for cars. And our point of departure is something else.

The underlying growth assumption for zero-alternatives used in impact assessments thus presupposed capacity expansions as the appropriate way of dealing with mounting congestion problems. This is quite problematic, since it severely hampers the ability of stakeholders that value other solutions than simply increasing capacity, e.g. transit orientated development or travel demand management. Since the impact assessments are mandatory and projects are often checked against a given benchmark of performance measures, these assumptions become important in determining which solutions to consider. In other words, the assumptions embedded in the appraisal tools favour capacity expansion over other approaches. This is somewhat unsurprising given the history of demand modelling (cf. section 2.5), as most contemporary transport models build on the original premise of facilitating motorway capacity for continuously rising levels of travel demand. However, facilitating sufficient motorway capacity is not the only (and arguably not even the main) challenge of contemporary transport planning, especially so in congested urban areas.

Overall, informants identified various known effects that were neglected in modelling, but which could explain inaccuracy in the form of both bias and imprecision. The most important effects of this seem to be a general neglect of induced demand as well as poor modelling tools for public and non-motorized transport modes. Several informants pointed to the considerable resources necessary for upgrading models as an explanation for why important effects are continually neglected.

Development

Another issue frequently mentioned by informants was that development of various key determinants often doesn't correspond with expectations. This was also the explanation that received the most support in the survey among stakeholders (cf. subsection 7.2.3), but interview informants generally blur the boundaries between different types of explanations. This means that the development concept represents not only unexpected development for external variables, but also design changes and lack of coordination with other planning projects – both land use and infrastructure - that impact the development of traffic on the project being assessed.

One local politician discussed how some of the underlying assumptions for a forecast are obviously out of touch with the most likely development, due to lack of coordination between land use and transport planners:

I think one of the things that limit models is concerning urban development. Once you've made an assessment... I mean, particularly for the alternative to the west, there was urban development plans in an area that – depending on the rise in sea level – would be completely under water. You have specified that there is going to be urban development here. And there has been a lot of debate about what is going to happen out there. I don't know how we can sell the lots at such high prices, because it is all below sea level if it comes to that. When these concerns were raised – why do we have urban development plans in this area – then they were taken out. But it was only taken out of the political decision; not in relation to the model [...] We are getting a new hospital out east, which you would assume to generate a lot of urban development. This might actually lead to a need for discussing new alignments. But these things are ignored. The model that was made – the numbers we have received – focuses on the urban development out west: the plans the city council already decided they didn't want to go forward with. But the assumptions were kept for the assessment anyway.

The decisions on land use planning are completely external to the model, but obviously carry a lot of influence on how much travel demand that is likely to materialize in different areas. A scenario needs to be specified for how the land use is likely to develop, but when the specification is based on plans that are already abandoned in political negotiations at the time of making the forecasting, it should be obvious that the validity of the forecasts suffer accordingly. These uncertainties cannot solely be blamed on the consultants, since the land use planning is ultimately a political decision that might not always be communicated explicitly. If a consultant becomes aware of changes to the framework that have important consequences to the validity of appraisal, there is obviously some responsibility of informing clients of this. It might not always be obvious that an informal political decision has been taken on these matters, and even if such a decision becomes explicit, this might happen well after the consultant has completed the contract. In these cases it becomes somewhat blurred where the responsibility for not adjusting the appraisal should be placed. However, many informants gave the impression that regardless of whether a consultant had objections to the validity of assumptions, the client always had the final call in these matters. One modeller, who had previously worked as both a consultant and researcher, explained how the consultant does not engage with such matters at all in his experience (emphasis in original):

This is never up to the consultant. Not in my world. Before I become a researcher I worked as a consultant, and the only things I did to make money was to use traffic models and build traffic models. We never made the decisions on what assumptions to use. That would be very dangerous... and there is no reason to do it, because that is what the local authorities are working with. I mean, if a municipality wants to build a tunnel... then the municipality needs to provide expectations for growth; car ownership, population, workplaces. It would be lunacy to expect a small consultancy firm to figure out all of these things. They have no way of doing that.

A private consultant supports this point, and makes an example of how some of the most important variables are typically outside the consultants' area of influence:

We ended up with a model that produced completely wrong forecasts for trucks when the bridge opened five or ten years later. But it was pretty accurate in terms of passenger travel. One of the reasons – and we did some internal notes on this – was caused by the information we received from freight forwarders. They do most of the freight transport for trucks in this corridor and apparently received hidden discounts from the ferry services. So the prices we got were typically 25% too high. We used these high rates for modelling the traffic on the bridge. It turned out to be much lower in reality, because these hidden discounts made it very feasible to continue using ferries.

Both of these citations illustrate the illusion of objectivity in transport project appraisal, since any model is only as good as the input it receives and the available input is often determined by people with a vested interest in the outcome of the forecast. Since many of the most important external variables are decided by policy makers or based on poor data quality, this introduces great levels of uncertainty to demand forecasts, since these are typically based on only a single scenario for the development of these variables. One planner described the problem as follows:

We don't know how the development is going to look like. I mean, very often the politicians will select the projected development that gets you the decision you want. And that is the political reality, right? That's not because transport models are bad. That's because we don't know if demand is rising. If we take this bridge project we made, then we said: "If the traffic growth is like this and this, then traffic on the existing network will grind to a halt this far into the future." And then people can say that they want to solve it like this and this. That's the best solution. Because you cannot really use the forecast for much. You simply can't. What is the traffic going to look like in ten years? We don't know what to answer. But you can use it – the assumptions – to say: "If we have 2% annual growth, then it looks like this." If it only grows with 1.5%... well, no one can say which it'll be [...] We tried to say: "If we have 3-4 alternatives, how does it look if the traffic growth hits different levels?" And how do they compare to each other; where does it start going wrong. And that [policy makers] can use to make decisions.

The main point here seems to be the importance of using different scenarios for the development of travel demand, so the assessment becomes relative rather than absolute. It might not be possible to predict actual development, but if one alternative ranks higher than others in the majority of

scenarios, then the actual development is not as crucial. This obviously reduces the impact of forecasting inaccuracy, but it also makes it difficult to compare conceptually different alternatives, if the uncertainty regarding development does not carry equal impact among these alternatives. For example, if a transport model is not able to reflect changes in parking policy, a relative comparison between a parking policy intervention and a capacity expansion would make little sense. A reduction of downtown parking spaces would certainly impact expected travel demand, but if the relative comparison is based on different growth scenarios that are equal among each alternative, then this effect would not be captured. This approach is therefore mainly suited for choosing between conceptually similar projects, when a decision has already been made to pursue a specific type of solution.

Another thing that informants were quite explicit about was the perception that deviations between forecasted and actual traffic volumes resulting from design changes are not to be considered forecasting inaccuracies. A project manager in a private consultancy firm commented on the issue in the following manner:

If you decide to build a 6-lane bridge instead of a 4-lane bridge... well, that is not an estimation error. At best, it is an error resulting from lack of clarity in your framework. But politicians love to add fingerprints on all sorts of things. And policy isn't rational. So it is no rational process to predict how political decisions are going to unfold during the project. I think you should move in the other direction and say: if we have major decisions on a design change, we should take the consequences and prepare a full assessment of it. Then we can say whether it makes sense. If it does, we put it in. And then there is no bullshit. I mean, it's just a new project then, which we need to assess holistically [...] That's fair enough. Because, to the extent that you accept that a decision has a consequence... if you stand by your decision, then we implement the consequence. Then we have a new project that includes the consequence you just accepted. Then everyone is happy. Because you just assessed that this change has enough value for you to live with the consequence [...] Some of the projects we have assessed for the rail directorate have a final budget that is over ten times the initial budget. It's pointless to measure forecasting accuracy on these. Obviously this happens because they decided to add this and that while they were going to dig stuff up anyway.

In line with the assumptions for input variables, there seems to be a generally accepted perspective among consultants that they are hired to give their best estimate based on a set of specifications. The specifications are left up to policy makers to decide upon while the consultants are mainly concerned with reliability of their models. But even the assumptions within these models might be wrong, and it can often be difficult to critique the result of an assessment when there are so many different locations of uncertainty to account for. A planner in the national rail directorate commented on the frustration that can come from knowing that a forecast is clearly optimistic, but without being able to specify where the cause of error might be found:

When I saw the forecast it was obvious that something was wrong. There was another forecast based on an older model, which produced considerably lower passenger estimates that seemed more realistic to me. Anyway, it was decided to use the newest

model, and that was it [...] I think it was obvious. The line had 25,000 passengers or something like that. An extension and a few extra stations to improve connectivity would result in a doubling of that amount in my opinion, perhaps a little more. Let's say 60,000 passengers. Then we have a model that says 100,000, so I think it was obvious that there was an error [...] The model determined the outcome since that was the one we had decided to go with. It was the most sophisticated model we had [...] Now the project is finished and we ended up with roughly the amount of passengers I expected. But I haven't seen a single critical comment about why we expected with 100,000. We joke about it internally, but that's it.

Two important notes should be made in regard to this citation. First, several assessments are typically made during the initial phases of a project. These might facilitate a learning process among stakeholders when they are used to explore different assumptions. However, when only one of these is presented in the decision support documents and no convincing argument explains how it was chosen, this obviously serves to mask a lot of the uncertainty regarding the results. Second, a newer, more sophisticated model is no guarantee for more accurate results. This is one of the typical arguments directed at the cause of observed forecasting accuracy being mainly political, since technical advances ought to have improved modelling accuracy. However, since models are built for different purposes and require access to high quality data, a more sophisticated model is no guarantee for improved results. In fact, if the necessary data is not available, a more sophisticated model is only likely to introduce further uncertainty. This is one of the fundamental rules of statistics; that 1a result does not become more robust simply by increasing the number of uncertain explanatory variables. A private consultant also expressed frustration at the lack of attention to specified assumptions when comparing expected and actual development⁶¹:

We made different model runs [...] Some with 50% urban development in the plan area, some with 100%... stuff like that. What hasn't been caught up by the media is that it was the full development scenario that was 70 million annual passengers [...] But politicians and journalists don't bother this. They only look at whether it carries 70 million passengers or not, and it only carries 40 million.

In the case presented here, the informant reiterates the fact that multiple assessments are constructed based on different scenarios for possible development, but that the main problem comes when a single scenario is presented as the only narrative. It might very well have been the case that a full development scenario would have resulted in 70 million passengers, but this scenario is not very relevant for the ex-post comparison when more suitable scenarios were in fact prepared. Once again, the problem of forecasting inaccuracy seems to be largely tied to unwillingness in accepting different possible outcomes that all need to be considered in detail.

⁶¹ The informant never addresses this issue explicitly, but the planned urban development was still not completed when the project opened, thus making the comparison with a full urban development scenario an unfair assessment of the forecasting accuracy of demand.

Another private consultant gives his perspective on why many rail projects fail to meet expectations due to lack of coordination between different projects, as well as why there might not be a high correlation between cost overruns and passenger shortfalls:

The data we use for price estimations is too old while the actual prices are heading towards the moon. This means you carve out the budget while the project is still being designed. It is a major source of error that I have never seen handled anywhere [...] We run back through the scope and reduce the project until it fits the budget. What typically happens is that you cut away stuff you can't really do without. Then you have a follow-up project to patch some of that stuff, which is typically more expensive than doing it the first time around. I mean, if you build a metro system and you want an underground depot... well, there is no money for that in the budget. But if you come back in 10 years and say: now we need the depot, otherwise we can't fulfil our performance requirements. Then it costs X times more to build the depot because you need change all sorts of stuff and it ends up becoming super complicated. If you had just dug the hole while you were there in first place, the costs would have been much smaller. Another option is that you start saving on components that results in higher maintenance costs. You might also end up a poor work environment because the available space gets too cramped. Your signal cabin becomes too small for example, so the next time you have to expand the facility you don't have room for a new one. I mean, you constantly sub-optimize to stay within the fucking budget. And that causes a shitty project. But all the rules are obeyed. Everyone is happy [...] The only problem is that the project is a piece of crap. But who cares? [...] Risks that hit your purpose are the worst. If you don't reach your purpose you need to initiate additional works. That increases duration, which increases costs. So a budget isn't the way to control uncertainty. You control uncertainty with a purpose and a schedule. All this focus on budget control; forget it. That doesn't control projects. It's just numbers.

This rather long citation displays a typical dilemma that arises in project management when faced with unexpected problems. Either you handle the problems by controlling purpose, duration, or budget. If you still want the project to fulfil its intended purpose, you need to adjust the budget and increase funding. However, since project management is mainly focused on controlling costs, you end up adjusting the purpose and thereby taking away functionality. In the end these adjustments result in a worse project than the one you based your assessments on, and consequently it likely ends up attracting fewer passengers⁶². This means that cost overruns and passenger shortfalls might in fact be highly interrelated, but through mechanisms that are not revealed when simply comparing the forecasting inaccuracies of these items across a sample of completed projects (cf. subsection 6.2.4).

The issue of coordination also receives attention from a local politician, who addresses the problem of public transport planning often needing to cater for car drivers to avoid losing political support:

⁶² An example of this is the Midland Metro in the United Kingdom, where several stations were scrapped to keep the project on budget. When the project opened, ridership was 38% below forecasts (UNITE 2011a).

When we have implemented the rail project we can always reduce the road capacity. You can't start by cutting away road capacity in advance. Then you get all the car organisations against you. We did this intentionally – continued with a 4 lane road – because if you do something else people will say no. They know what they've got. When people start using the rail system we can always cut some road capacity. Or do like the US, where you need to carpool to go past a certain point [...] You can't impose restrictions until you have an alternative. The main reason is to ensure that you don't end up facing unnecessary political resistance.

This citation is more a reflection of how decisions are taken in a politicised environment, where the solution you want is not always the solution you are able to get; at least not the first time around. The result is a final vision for the project that does not correspond entirely with what the conditions in place when it opens. In time the project might be able to fulfil the ridership projections, but this requires some political motivation to follow through with the restrictions on car use that formed an important part of the original vision.

Overall, informants identified a number of issues related to the development of external variables that might cause forecasts to become inaccurate. Apart from a general uncertainty related to growth trends for economic development, several informants pointed out that data quality is relatively poor and often outdated. Furthermore, many of the most important determinants for travel demand are based on political decisions that are not always communicated to modellers, which make it difficult to account for them in the decision support prepared for policy makers.

Politics

One issue that every informant addressed in relation to explaining forecasting inaccuracy was politics. However, politics as an explanatory category should not be confused with the incentive structures for deceptive behaviour that is often touted by proponents of strategic misrepresentation as the primary explanation of forecasting accuracy. Some informants mention these incentive structures, but the politics category relates more generally to the complex and often irrational political system, that increases uncertainty regarding many of the important variables used to determine future travel demand. We have already seen some examples of this in a few of the citations used to describe the previous two categories (neglected effects and development).

Informants who worked mainly with the technical aspects took great offense to any accusations of deliberately manipulated figures as a result of political pressure. One modeller claimed that what was perceived as strategic behaviour was often just a matter of updated models as a result of new information or access to better data:

Based on the data we have had available we estimated the connection between income and car ownership. It's possible that if we re-estimated the model - with the knowledge from the newest national transportation survey – we would get another relation between the two variables. I don't know. But you can't tamper with the model. I mean, the coefficient that is produced by the model... that's the coefficient, right? Bam, finished. Then you can evaluate whether you believe in it or not. If you don't believe in it, then you shouldn't use this variable. But if it has the right sign and the right elasticity –

compared with similar projects – then you can't say: "I don't want to use this". Then you have to use it [...] A representative from a public transport company explained the problem at a conference some time ago [...] Because she thought: "my God, why don't we do this more often? Why don't we make the same calculations with different versions of the model?" Then we could say: "We aren't cheating with the figures. We are just able to provide better information as we improve our models." [...] In my 20 years working with this – either as a researcher or consultant, or now mostly as a civil servant – I've never been in a group of people who would say: "Let's make this look a little better." I've never experienced that situation. Why would I risk my job? For all I care they can build 15 motorways between these two cities; I just don't give a damn. But the calculations need to be correct.

It was fairly obvious from the interview that this informant was furious about accusations of fraud being a major source of inaccuracy. He goes on to express his frustration about the accusations towards the profession:

I don't mind people asking questions: "Is it really realistic to have this many passengers?" That's no problem. But it would be nice if these people were then able to tell me what was wrong. It's like – when you are so engaged in what you are writing, you don't spot all the spelling mistakes – it's the same with the models I build. Sometimes you are just staring at that program and you don't spot a huge error. Others might spot it immediately, but I can't see it. So it would be nice with a capacity like Flyvbjerg coming in and saying: "Hey, this is overestimated, there's an error here!" Then I would say: "Ah, right". But we have quality checks and assumptions from statistics bureaus and we have estimated the importance of this and that, right? And the results are in line with international projects of this kind, right? I also have a budget and a deadline, so what do they expect? What is it they want?

It is important to note that the citation solely relates to direct tampering with the model parameters. Most informants seemed to agree direct tampering with the model parameters was unlikely to be a major concern, since those able to do it would risk a lot by attempting such an obvious swindle, while those who might gain from it had access to much easier ways of adjusting assessments if they really wished to manipulate figures, e.g. by specifying favourable conditions or selectively reporting appraisal results. Many respondents also pointed out that economic performance measures are typically downplayed rather than manipulated, e.g. if the project has political support but a low internal rate of return. A modeller describes the situation from his perspective:

We are doing a scoping of a motorway here. But you'll read in the paper that it is already decided [...] It's part of a larger pool of political decisions where you have to admit that we don't need any modelling. The politicians decided these projects already. It's like this other big project down south. There's no cost benefit in it. But then the minister will say: "It's good for the country, good for the export, lots of jobs." And that's fair enough. That's their job, right?

All informants, politicians and otherwise, had an almost unanimous acceptance of the fact that model-based decision support was only a small part of a much larger set of priorities that needed to be taken into account when evaluating transport policies. But all of them denied that manipulation was a large part of the process. A local politician explains it from a decision-making perspective:

Clearly there are some expectations. And you need some data to show how the recent development has been in this area... which will also indicate how the things will continue depending on what alternative is chosen. I don't think civil servants approach consultants and tell them: "You need to work towards this decision." There's a high degree of professionalism on both parts [...] Politically, you have long had an opinion on what solution you preferred. Then you say: "Now we have these figures". And then you use them based on your standpoint [...] You can relate to numbers. You can't relate to a consequence about something that might have a negative influence on something. Something that is not very measurable is difficult compared to numbers. And that's the point; if you want to use numbers you need to know what they mean, how they are produced.

A researcher, who works part time as a consultant, supports the sentiment and goes on to explain how the knowledge produced in assessments can lead to design changes:

The political agenda wasn't that they needed this project. It was rather to send a signal of a community that experienced growth. That's often where it starts. And that's not really surprising, that's still the case most places. But, at that point we did have a lot of growth. All the suburbs had been booming, and a lot of people had moved to the area [...] I have no knowledge of anyone deliberately manipulating anything. I simply have no evidence that enables me to relate to that [...] Of course it is blurred, because a lot of things never reach the public. All these assumptions... I mean, a lot of times I've experienced that a client didn't like the results.[...] Then you are usually asked to have a discussion about whether there is something that can be changed. Are there assumptions that should be different? And that might produce a different result... but it's not a wrong result. I mean, we had an urban rail project where we assumed it to run 80 km/h. But it turned out it wasn't very competitive. And then we said: "Well, it's possible to get running stock that goes up to 120 km/h for this type of project... so we can assume a higher speed." This requires the tracks to be fenced. But there's nothing wrong in making that change. It just results in some barriers along these routes [...] You try to change the conditions in order to make the project more competitive. There's nothing wrong with that. You just agree that the design can be changed.

Apart from supporting the argument of politicians using information selectively rather than seeking to manipulate it, this citation also indicates that forecasting inaccuracies as a result of design changes might actually be caused by the forecasts themselves. This adds to the complexity of performing ex-post evaluations, and calls into question whether increasing costs as a result of scope creep should always be considered problematic. It might in fact be a result of attempts at improving overall project efficiency. Another possibility is that it is simply caused by error. One project manager in the rail directorate dismissed intentional deceit as a main explanation and looked to incompetence instead:

There was this philosopher once – William of Ockham – who established Ockham’s razor, to clean up all the nonsense that circulated at the time. He basically said: “If you have to explain something... the best explanation is probably the simplest.” And then I read about someone who called it Fuckham’s Razor, which basically states that... well, people often see conspiracies – clever alliances to pull things in this or that direction. It’s probably a conspiracy when something looks out of place. But then Fuckham’s Razor says: “Listen; if you can explain a process by someone fucking up somewhere, then you can forget about conspiracies.” I think that applies very well in these cases.

This issue is often somewhat tied to the earlier notions of resources, since fewer resources obviously reduce the quality assurance tied to a forecast. Not only because it typically results in less sophisticated models, but also because consultants have less time to validate model results. One civil servant addressed this issue in detail:

Let’s just say I have spent 100 hours on this project - realistically it’s many hundred hours, but let’s use a simple example – then I have spent at least 85 of those learning, controlling, correcting, and asking for new calculations from consultants. But why is that? Two things. First, the consultants didn’t have the knowledge I expected. Second, the modelling procedure was of lower quality than I expected [...] Unfortunately it is my experience that you often have too little time and not enough money. You have to cut the amount of scenarios or the number of model runs. If the consultants do not have the capacity to weed out errors and be critical towards results... well, then we’re in trouble [...] I don’t know what it takes really - maybe fewer projects.

Another civil servant explains how these errors can be problematic if they are discovered too late, since the results might already have been used to argue in favour of a project:

I don’t think there is a pressure to produce certain results. Not in the initial phases anyway. We can sometimes experience another kind of pressure if you have done some stuff early on, right? Everyone is happy about it: “Hey, what an exciting and interesting result! We need to take this further up the system” Then later you figure out it was a mistake. I mean, particularly if it’s an error. An embarrassing error: “The consultant miscalculated. It’s not 1.3 billion, its 1.7 billion - because we made a mistake - no, it’s not because we have new information. It was just an error.” I mean, especially if it has already been flashed a bit.

He goes on to explain that it works from a top-down perspective as well:

You don’t just keep throwing candy at a project to make sure you have covered everything. Then a retarded project manager comes along. We need to add 15% then, because he can’t keep things within budget, right? So the budget just increases endlessly if we do that. I mean, if you have incompetent people then it’s bound to fail. You need a scandal then. But that’s fine; it’s how you keep your organization lean.

The point here is that confirmation bias might cause proponents of a problem to run with a result before conducting proper quality assurance. When the mistake is realized there might be an indirect

pressure to keep the knowledge contained. Not necessarily in an attempt to deliberately boost the assessment of the project, but rather to avoid confronting your superiors with an embarrassing mistake you should probably have noticed earlier. Such incidents typically cause an outrage regarding professional integrity if the media catches on to it⁶³. Conversely, if no one ever takes responsibility for any mistakes, the environment doesn't foster any learning or incentive for improvement.

Resources and incompetence are thus often used to explain what outsiders perceive as manipulation. This doesn't mean that no politically motivated behaviours can be identified, but rather than intentional deceit they are described as 'rules of the game' that everyone are more or less away of and accept. Informants generally speak of such behaviour in relation to what they consider cunning tactics of gaining the most of a situation in the various forums that decisions take place in. One project manager in a private consultancy firm offers the following explanation for how strategic behaviour can influence the accuracy of forecasts, albeit in a manner that is not necessarily deceptive or intended to misrepresent facts⁶⁴:

Informant: During the process you have hearing phases. Hearings always result in additions. The transport authority made an interesting model at some point. I didn't think of that in relation to this earlier. They had a model with three levels as I recall it. They used it for a project some time ago, where level one is the minimal solution that can be approved. Formally that is; the one that fulfils every requirement, but nothing more. Level three is where you have all the nice-to-haves, the one everybody can agree upon. Level two is the qualified proposal. Let's say you have a known budget, since you have been through the scoping phase. If you went into a hearing, which level would you present?

Me: With the known budget?

Informant: Yes. You have a project that needs to pass a hearing. You made some early cost estimates. Let's just say the figure is somewhere around here (marks an x at level one). That's your budget.

Me: If the purpose is approval I would present this one (points to level three).

Informant: But you don't have the budget for that one (laughs)! But you can get it. When the hearing is over, then the number has become $x + y$. But what if you don't get all the extra money? Then the project dies... because the public demands bike parking every 50 metres. Or they want bike paths all along the project. Or parking lots and bus terminals. Or compensation for this and that. You name it. The creativity never stops, right? Everyone within 14 kilometres wants to be expropriated, because... all kinds of stuff really. Noise walls. Giant noise walls everywhere. It's endless. And it's the planning

⁶³ See e.g. Marfelt (1996, 2010) for Danish examples.

⁶⁴ I offer this citation as a correspondence between the informant and myself rather than an shorter edit, since the format was very interactive with plenty use of whiteboards for this interview. I feel that it is difficult to do the justice to the informant's message without including this interaction.

law that dictates this process. It ensures some sluggishness and consensus building in the system. But it also makes it incredibly difficult to propose something, because what the hell is the price going to be? I mean, you can practically guarantee that no matter what you present, that certainly won't be any cheaper than what you end up building. There will be some additions along the way. Then we say: "ok, fair enough. So these additions we expect, they must form a reasonable project." But if we present that instead, then we just have other additions. Because the mechanisms of a hearing requires that something more is added. The public needs to achieve something extra compared to what was proposed. So the transport authority's thesis was that you propose the minimum solution. And then you set aside a reserve for what is considered reasonable additions. So in practice you say; this isn't our real project (points to level one). This over here; this is our real project (points to level two). This thing here; this is our uncertainty (draws a line from level one to level two).

This rather long citation presents an interesting perspective that runs somewhat counter to the typical criticism of strategic behaviour. Rather than presenting an obstacle to democratic decision processes, the informant considers the behaviour a result of democratic decision process. Not to exploit the process to obtain an unfair advantage in influence over which decisions are taken, but simply to avoid that projects escalate beyond control and end up being abandoned. Conversely, neglecting to account for demands from the public runs the risk of increasing costs due to protests. Newbury Bypass in the UK is probably the most famous example of this, where a pro forma hearing process resulted in massive protests during construction. The additional costs incurred by protest action made up more than one third of the total scheme costs (UNITE 2011a).

Overall, informants identified a number of politically motivated incentives that could influence the accuracy of demand forecasts. However, rather than focusing on manipulated figures as a way of misrepresenting forecasts, informants pointed to mistakes being subdued to avoid appearing incompetent as well as a general acknowledgement of scope-creep regardless of how a project is specified. The latter point runs counter to traditional theories of strategic misrepresentation: full disclosure is avoided as a precaution to avoid rampant scope-creep rather than as a way of facilitating later scope-creep.

7.3.2 Communication

Most informants brought up problems of communication as a major issue in relation to uncertainty during interviews. While the relationship among stakeholders was a minor agenda point in the loosely structured interview guide, respondents were never made aware of this and often addressed the issue before I had a chance to bring it up myself. All informants viewed model-based decision support as only a part of a larger set of tools for communicating expectations and uncertainties, albeit a very crucial part that was often given an unsatisfyingly lack of attention in terms of critical scrutiny and clarification.

Uncertainty

Considering the context of the interviews it is probably of little surprise that communication of uncertainties in relation to model-based decision support was one of the most addressed topics by informants. Where many of the available decision support documents and survey respondents

tended to focus on uncertainties located in the models or the external variables, project managers and politicians spent more time highlighting the enormous uncertainties stemming from lack of clarity about how to define a project. One private consultant explained in the following way:

You need to consider if you want to accept additional costs in order to get a more effective solution. Decision makers just aren't always willing to make that call. The terms of the project need to be specified at a very early stage, because they are going to dimension the entire project. If you don't call the shots from the get-go... well then it comes back as nasty surprises later on, or you never consider the best solution. Because you made a set of assumptions that no one has seen or understood. It's just... I mean, as an engineer you need these assumptions to work with. If you haven't discussed them with anyone, no actual decisions are taken. Then it just lies there in the stomach of the model. So these things need to be pulled to the front of the planning reference [...] Let's say you need to pull new cables; often you won't know where the old stuff is buried. If you want to figure that out you have to dig up the old cables anyway. Is there room for new ones in the old pit, and will it work with the old configuration? If you have already dug it all up, why not rewire the entire thing? Instead of spending resources on mapping all the old shit and do some half-hearted solution. So where's the interface? How much is in the project and how much isn't? These investigations need to happen early on, and result in a set of decisions based on principal or strategic prioritizations regarding how you want to do things. A lot of this stuff is cut away to save money... let's just get started, right? We need some results right away. And when that happens, you carry all this uncertainty around.

The type of uncertainty addressed here relates more to the context than the model if we frame it in the location dimension of uncertainty (cf. section 2.3). As was indicated by the last citation under politics in subsection 7.3.1, the vague definitions might even be intentional in some cases, to allow expected design changes to be implemented at a later stage. Whatever the reason, lack of clear project definitions is a major source of uncertainty, and if the framework is open for interpretation, the consultant will either ask for clarification or make a set of assumptions. The latter situation is probably the more likely of the two, since clarification by the client risk resulting in an expanded scope of inquiry with no corresponding increase in budget. This is necessary for any useful forecasts to be produced and not in itself a problem, but it requires that all the minor decisions and assumptions that end up embedded in a forecast and stored only as tacit knowledge are somehow made explicit to other stakeholders; especially policy makers. A modeller claims that this is certainly not the strong suit among traffic engineers, which is one of the typical educational backgrounds for people with the producing demand forecasts:

There are two things engineers are generally poor at; communicating results and creating awareness. The things we're working with... that's something you can use, right? Engineers suck at communicating that. So these notes that are produced on calculations are just numbers and tables. But it rarely turns out the way it was presented in the table. It's a classic; in the eyes of an engineer, communication is something politicians and civil servants are concerned with, because...we just deliver results, and

then it's up to you to figure out whether we have done enough [...] In a long process of building and applying a model, the quality assurance and documentation is something that happens near the end. The money has all been spent a long time ago at this point... due to model errors and stuff like that. That last part about communication and documentation is clearly the place in the process where quality suffers the most [...] I worked on this large project a while ago, and no matter how much money I got... I was just as awful at documenting it and all that stuff. A traditional model builder is so focused on making sure that things fit together. You check everything again and again and re-estimate the model a million times... and this variable here, it needs to be in there. Is it logarithmic or just linear? There's all kinds of that stuff. You just... whatever. We'll spend the entire budget, no problem. If I was given triple the budget I would spend that too and produce just as poor documentation.

According to this informant, the lack of quality in documentation shouldn't necessarily be associated with poor quality of modelling. Communication and documentation simply isn't a key priority among people who produce the forecasts, and unless specific demands are made for increased communication of uncertainties this is unlikely to happen. Modellers are simply much more concerned with ensuring the quality of the models than informing others of the technical details and the implications for interpreting results. A researcher goes on to explain that this makes it very difficult to do any substantial peer reviewing of results:

I'm not always satisfied with the way consultants are documenting the assumptions or the input to their analyses. They're often writing a little bit too fast about all these assumptions and all these things that make it very difficult for an outside reviewer to see what the assumptions actually were. Was it, as you say, that they estimated a very fast economic development, or was it that they expected the population to go down, or why did they come up with these results, what petrol price did they assume for the future and things like that.

A private consultant explains that decision makers typically won't ask for such information either, but rather push for it to be excluded:

People can only cope with one alternative in their heads at a time. You aren't really grasping the full span of uncertainty that way. You only focus on that one technical solution. And that's a mistake; you don't zoom out to see the whole picture. But, I mean... the head of a politician can't deal with estimates of $\pm 50\%$ or $\pm 100\%$, which is often the case out here. It could be $+400\%$ and -50% for that matter. But they can't relate to that at all. "But what does it cost?" We don't know, that's the whole point! But they want a number and then they just throw in a span. "What uncertainty do we usually have on this sort of stuff? 30% ? Bam, done." That's not an estimate... that's just babble. I don't know what the answer is [...] There's no point in trying to measure it with such accuracy. It's unrealistic. I mean, if we have a hundred possible futures, how are we going to select the right one?

The point here is that uncertainties are often recognized by the people working on the technical aspects, but there is little incentive to report it since no one asks for it. In fact, policy makers often ask for the exact opposite, and expect engineers to produce as exact figures as possible. A planner supports this perspective with an example from a previous project:

I can think of a few places where we have done it. We're doing a large rail project now where we spent a lot of time describing uncertainties, what we think about all the involved parameters and stuff like that. We did it for the strategic analysis and sort of hinted that the results are kind of a big span. But ... it was too complicated for anyone to understand it. "Oh jeez, you come here with a ton of results? Which one is the right one then? Do we go for the one in the middle, or?" It's a difficult message to communicate that there's a spread and not just a mean [...] It's too difficult to understand for them: "What do we need these nuances for? Do we build it or not?" At the end of the day we need to recommend something. "Just give it your best shot." [...] They don't know what to do with it. Or why they need it: "What is this for? Does that mean it's a bad idea?" Decision under uncertainty is an unknown concept.

This citation seems to suggest that decision makers are not always comfortable with taking responsibility of their decisions when the consequences of them are uncertain. This is understandable to a certain degree, but it is also the very reason we have elected public officials, after all. Their primary function is to make difficult decisions in matters that are too complex to derive an optimal solution scientifically. A modeller supports the sentiment of poor documentation of uncertainty, and mentions that part of the problem is also imposed by the structure of the tools used for subsequent impact appraisals:

When we do the socioeconomic assessment we have different levels of costs. A mean value and 10% cheaper – for whatever reason, no one ever believes that – as well as 30% over budget. Stuff like that. But the benefits you get – the time savings from the transport model – you'll just see a single figure there. There is no $\pm 30\%$ for that. So I'm asking: "Why can the budget estimates be uncertain but not the transport models?" That's against all reason [...] In the model you can play with cost intervals, but the benefits have to be a point estimate. Maybe I shouldn't bark too loud, because the ministry of transport might drop by tomorrow and say: "Can't you make these intervals?" (laughs). But what should they be then? $\pm 100\%$? [...] Most people expect the model to throw out a single figure. But that's not the case. It's a figure plus an interval... the standard deviation. We just forget about that part. We look at the mean, right? The deviation... who cares? Moshe⁶⁵ claims that the standard deviation is like $\pm 100\%$. We might get 100,000 passengers, but we also might get... well not zero, but 50,000. Or 200,000.

The comment here seems to relate to a more fundamental structural problem. Even if vast uncertainties in demand forecasts were in fact communicated, they would probably still be excluded from the final impact assessments, since these require relatively exact estimates of costs and

⁶⁵ Moshe E. Ben-Akiva, professor at Massachusetts Institute of Technology.

benefits. An alternative approach is of course to prepare multiple scenarios for these point estimates, but this can become fairly resource consuming since it would likely require the entire appraisal to be completed for each scenario. As an example, the Nordhavnsvej case presented in subsection 6.3.2 illustrated how benefits cannot simply be scaled up in accordance with new demand estimates, since the effect of additional traffic is neither linear nor mono-directional in relation to travel time savings. Even if you somehow managed to include these uncertainties in the documentation, the information might never reach decision makers and have an actual impact. One parliamentary politician explains it as follows:

I don't think many politicians – not in the transport committee or elsewhere – read those assessments at all. I don't think so... you read a summary. It's in the annotations for the construction law when it is proposed, and you read that instead [...] You don't have the time for it. Time-wise it is impossible for politicians to read anything beyond the stuff that requires special attention. Luckily you have environmental people and researchers that do read this sort of stuff and will create awareness if there are problems. Luckily someone reads it; but not necessarily politicians.

A local politician agrees when commenting on an impact appraisal for a new project:

If you were to ask the city council, my guess is that no one has read this thing. And it's not a unique case; it happens all over the place. Just look at the new local plan we put in hearing. People make grand speeches about how good it is, but don't try to tell me they have read the 6,500 pages that it consists of [...] I doubt anyone in the city council has doubts about their position on this matter.

This point is perhaps obvious but at the same time quite important for the communication and treatment of uncertainties. When politicians decide what goes into an assessment, both directly and indirectly, they also control the information that is available to the public. But if politicians do not ask for comprehensive documentation of assumptions and uncertainties because they have already formed an opinion on the matter, it is difficult for the public to act as the necessary watch dog that the parliamentary politician above says he relies on. He goes on to explain how this often impacts the political discussion when uncertainties are hidden:

I mean, we're aware that these numbers aren't exact, because we know that you can't forecast with such precision... but I think a span of $\pm 25\%$ would surprise most politicians [...] It's more like: "Well, there are some numbers here, so it has to be true." I've been in debates where politicians have argued in a manner that presents these numbers as... well, if not with pinpoint accuracy, then at least very close to it. So there's no doubt that the construction law ought to include annotations regarding the uncertainty of forecasts. People might not notice it, but at least it ought to be in there [...] It's a bit like when economists do their projections. It usually says there's a range of uncertainties. That it's the best result given our available knowledge. And that's fair enough. The problem arises when it is taken as a fact in politics [...] That's why assumptions are so important. But before you know it, you are knee-deep in a philosophical critique of the processes we engage in. We don't have the bloody time for that.

The perception that the vast majority of information in decision support documents never reaches the decision makers is shared by all informants, as is the perception that this creates a lot of leeway in terms of interpretation. One private consultant airs his frustration about being misinterpreted, but claims that this is part of the game:

Perhaps it shows our inability to communicate things clearly, so there is little room for misunderstanding. But it's difficult. We always try... in these reports we always have 3-4 pages of executive summary aimed at the minister or whoever it is that only has a couple of minutes to read all of this. There it needs to say what we really mean. But then you have a civil servant who goes: "You know what, here it says so and so, which actually means this and that". But that's not our fault – I mean, we can't do much about that – and we don't get involved in the political debate. We stay out of that. Then you can accuse us of being a bunch of sissies, but we just won't do it.

He goes on the state to give an example of how misinterpreted statements can create a 'once bitten, twice shy' mentality among consultants, who rely on their perceived neutrality:

I was asked by the ministry a while ago to give an interview to a journalist regarding different public transport alternatives, and he wrote an article where I agreed with the content. But he forgot to tell me what the heading was going to be, which resulted in a phone siege of people asking if I really meant what it said in the paper. I had to go out and say: "I didn't write the bloody heading!" I was really hurt, because it said that I thought rail projects were a waste of money or something like that [...] I told him: "I'm done giving interviews". It was picked up by a pro-rail NGO, but I refused to make further comments [...] If you read the whole article you'll see I never express an opinion in favour of any specific solution, but that damn heading...

A consultant, who works part time as a researcher, supports the perspective and explains how the roles of consultant and researcher are quite different due to different degrees in what you are able to debate publically:

Uncertainties are complex material. It's difficult to communicate uncertainty in a good way. I don't think there's a problem in communicating them between professionals. People in the transport authorities know what we are talking about when we speak of uncertainties. But when you approach people who are not professionals – politicians – these are different cases, I'm afraid [...] Other people have to raise their voice. It was not an issue for me to do so. As a consultant you typically keep a low profile in these kinds of issues. Usually we don't get into a debate with the client. As a researcher – I work part time at the university – that is another issue. There I feel free to discuss things.

This citation once again highlights the importance of documentation, since peer review is only possible if assumptions, specifications, and data are made readily available. As has hopefully been evident throughout the present thesis and especially in section 5.1, this is certainly not the case for the vast majority of transport infrastructure projects in Scandinavia. The results presented in section 3.2 also indicate that this is a problem that extends globally.

Overall, informants brought up several problematic issues in the communication of uncertainty. The most important seem to be a lack of interest from policy makers, who generally expect results of technical analyses to be both objective and accurate. If results are complicated by too many conditionals or reported with large uncertainty spans, policy makers become unsure about how to use the complexity of information they are being presented with. In addition, there is a lack of incentive for consultants to engage in public debate over technical analyses, even if they consider the implications of results to be misunderstood or misrepresented. As a result of poor uncertainty documentation and reluctance of consultants to engage in debate, it becomes difficult for people outside the formal decision-making process to scrutinize the validity of impact appraisals.

Interpretation

The concept of interpretation has already surfaced in the previous discussion of uncertainty, but in relation to the evaluation of transport projects it extends far beyond the mere technical details of calculations presented in decision support documents. Interpretation of the purpose of planning, the role of different stakeholders, and the objective of various impact assessments were also frequent topics that informants brought up during interviews. There are sometimes radically different interpretations of these matters among stakeholders, but if such issues are not brought forth in communication it is of course difficult to reach consensus on the appropriate way to assess a project. As a result, it becomes difficult to reach consensus on an appropriate solution.

One private consultant was quite clear in his interpretation of this:

We have this internal thing. We all go: "But it's just a game!" I mean, it's just estimates... at the core, right? We don't know anything before we build it. Sure, you do your best. But it will never be a science. It's just estimates.

While this was obviously said in a humorous tone, it bears a lot of truth if we recall the previous discussion of inability or reluctance among decision makers in engaging with uncertainties. While often expressed in a more serious tone, the perception was shared by most informants, especially those working with the technical aspects of forecasts. Those who produce the results acknowledge a vast uncertainty, but those who use the results perceive them as being very precise. The latter statement is probably not the whole truth, since decision makers often do acknowledge various degrees of uncertainty. The perceived uncertainty among decision makers just seem to be considerably smaller than the perceived uncertainty among modellers, and is probably further reduced when the results enter a political debate. The consultant from the previous citation goes on to explain how inability to accept uncertainty can be a cause of inaccuracy, but that they shouldn't necessarily be interpreted as problematic:

What happens is that we end up presenting an average. That's the figure we believe in. It's the most likely number. But the most likely number is useless to politicians if they stop to think about it. A project almost always comes with a guarantee that it will be more expensive. It always creeps upward. And there's a good reason for that. I worked with – regarding time schedules – a sort of qualitative uncertainty evaluation. And it is simply human mechanisms that cause these things to creep upwards [...] If someone is supposed to finish around here and actually had the option of finishing early.... well, that

is never reported, because then he sits around and fiddles with it a bit longer. Polishes everything up a bit, right? And then he finishes exactly on time. So you never have the positive deviations. Everything is biased towards delays.

The point here is that the forecasts produced during the lifetime of a project can themselves influence the development and thus the validity of the forecast itself. If the timeframe is a week, chances are that roughly a week will be spent. Any unforeseen problems that affect the schedule result in delays. Meanwhile, any unforeseen benefits that affect the schedule result in more time spent on the same assignment until the week is up⁶⁶. The same is likely true for costs. If the budget is one million and the project can be completed for 900,000, the last 100,000 will likely be spent on small tweaks and upgrades that weren't necessarily part of the original project. After all, the money is available and it will result in a better project. The issue of interpretation then becomes relevant in the discussion of whether to consider a normal distribution of forecasting accuracy as a reasonable expectation. An alternative version is of course that the unspent resources are used to cover problems encountered elsewhere, but the net result is more or less the same.

A planner in the national road directorate addresses a more direct version of interpretation in relation to the results of impact assessments:

I don't think it matters a great deal. Even though politicians claim to be concerned with socioeconomic assessments, I'm not so sure that this is actually the case. What they care about is whether there are queues in the system. Then they think we need an expansion here. And... well, sometimes the calculations show that it is a good idea. And sometimes they show that it isn't. Then they might decide it anyway of course (laughs) [...] But that's also the way it should be.

A modeller in the same directorate supports this perspective, and claims that politicians have long ago decided on the preferred conceptual solution when they approach consultants for advice:

Some politicians say: "We want asphalt." Then they ask the road directorate to find the optimal way of doing that [...] Then they take it to the transport committee, where it is approved or dismissed. After approval it goes to hearing [...] But the politicians don't ask us what course to chart; they read the situation through their own lens. Then they ask us to answer other kinds of questions. And then you answer those. But many politicians interpret accessibility and mobility as equalling more asphalt. That might be too simple a way to put it, but I think it's a pretty legitimate assessment.

This is in line with some of the discussion presented earlier, indicating that while politicians might be influenced by impact assessments they do not necessarily value impacts in the monetary terms of a cost benefit framework. Rather, they embed the knowledge obtained from the impact assessment in a more holistic assessment of the project in its broader societal context, which may or may not correspond with the results of a more technical impact assessment. However, there is ample room for confirmation bias to set in, since a particular type of solution is preferred already at the initial

⁶⁶ This phenomenon is perhaps more humorously known as Parkinson's Law, stating that "work expands so as to fill the time available for its completion" (Parkinson 1955).

assessment stages. A local politician supports this sentiment, and argues that politicians will use results that fit them, and otherwise look elsewhere for support to their established principles:

The problem is that you just focus on something else then. Traffic will get stuck or something like that. I mean, this new bridge we are discussing at the moment is really a great example. The claim is that we will see mounting congestion if we don't get a new connection. Traffic will grind to a halt and the only way to resolve the situation is to build a new bridge [...] The internal rate of return isn't even looked at. If you look back, it hasn't been part of the debate at all. If you tried to argue that this or that solution would give a better rate of return... well, there has been no focus on that, because it doesn't matter. In these cases politics becomes religion. And you find the numbers that match you principles. Perhaps if you could put something into the consequences – some of the softer values – for all the stuff going on out there [...] But that isn't done, and probably because it's difficult to measure.

This citation illustrates the common perspective among informants that impact assessments do not determine decisions. At best they support them, but often they merely justify them; a perspective much in line with the results from Figure 55., if no justification can be found, the results of the appraisal have little relevance as politicians have often decided on the desirability of a project well in advance. A private consultant goes on to explain that since many important political prioritisations are left out of the assessments, there is good reason for politicians to look for arguments in support of projects elsewhere:

You have these disaster projects that end up being a success. I mean, technically –in terms of the three parameters you always measure them against – they are really bad, but in practice they are actually quite good. Great Belt Bridge or Sidney Opera house for example. Classic examples like that. I think it's just a matter of too much tunnel vision regarding potential benefits. You pick something you can measure – the amount of cars or passenger miles in public transport – but you forget all the rest. I mean, self-identity, development of society, integration effects, and balance in relation to the larger European cities... there are so many other benefits than those passengers. I think you oversimplify your benefit framework. And then you miss your target. I think if you had a broader scope and made benefits more holistic in relation to society – and not just the transport effects – then you would get more valid evaluations [...] You ought to do ex-post evaluations on the business case for the Copenhagen metro. Why are people happy about it? Why are the perceived benefits so positive? People have forgotten what it cost, but it works like a charm.

This comment relates to a fundamental criticism of the core foundation of cost benefit analysis, or at least the way it is typically employed in practice in transport project evaluation. Such analysis attempts to value all costs and benefits in monetary terms in order to make a structured evaluation of projects that enables comparison across a range of alternatives. However, when only the benefit categories that are considered reasonable to measure are included, this skews the evaluation of projects with many benefits that fall into non-measurable categories. If these benefits were similar for all projects this would not be a problem, as the relative comparison would still be useful.

However, this is probably only the case for relatively simple projects, and in these cases there is still some inconsistency in comparison with the zero-alternative. One planner in a public transport company explains how this is a problem in relation to public transport systems, and uses the case of a planned urban rail system as an example:

We have local politicians here who are convinced that this is a good project. I have a difficult time connecting their values with the things that happen when we do these evaluations and calculations. Their basic perception of this is that this project is a necessity when they look 10, 20, or 30 years ahead. They view things in a longer timeframe so to speak [...] They don't believe it is possible to manage the demand with bus systems in the long run. So for them it's a conviction that we have reached a size that demands this type of system. If they look to international cities of comparable size, they see that they all struggle in providing good public transport with bus systems [...] They have already started adjusting all the surrounding land use plans to make urban development along the corridors. Their priorities are far into the future, and they don't give a damn about the socioeconomic evaluation [...] How do you reflect these things going on inside the head of politicians? I mean, 50 years ahead with a new urban structure, where they think we need a different transport system... it's extremely difficult to relate that to these stark economic evaluations [...] When we discuss it with the ministry they tell us to add these things in annotations to the socioeconomic assessment... to mention that this is part of an urban development strategy. But to the politicians that is the main priority. It's the number one thing on the list, right?

The point being presented here is that since the most important benefit categories to politicians aren't included in the assessment, the assessment carries little value as decision support and is thus ignored by them. Cost benefit analyses of transport infrastructure projects typically focus on transport effects such as time savings or revenue from ticket sales, but for the politicians the project is part of a much larger development strategy. This relates to the earlier comment from a private consultant about having the benefits as the main control feature rather than the costs. The project is a necessary piece of a larger puzzle aiming to enhance the long term mobility in an urban area, and whether it results in immediate improvements for travellers is not the primary concern. This planner from the public transport company goes on to argue that the cost benefit analysis is mainly of importance to local politicians in obtaining state funding, since they know the project is benchmarked based on its internal rate of return:

Of course it matters if we come and tell them: "Well, when we look at this it seems like it's going to be really expensive. And it's a lot of money out of your pockets for operating costs." Then it starts to hurt, right? I mean, do we replace the bus system then? Is it now we do that? [...] So my honest answer is that that when we bring in an assessment that says 2,3, or 5% internal rate of return... they don't care about that at all. Except... they know that in order to get state funding, it needs to look somewhat reasonable so the ministry of finance won't cut them off [...] Locally it doesn't matter. Our assessments are made because it's a requirement from the ministry. It's not a requirement from any local actors. Here they just care about operations and maintenance.

This citation is of course of some importance in relation to the political and economic incentive structures mentioned by several authors as a major explanation for forecasting inaccuracy. If local stakeholders are convinced about the success of a project regardless of the socioeconomic assessment, but consider a positive result a requirement for approval at the national level, this obviously creates an incentive to inflate benefits. However, this is too reductionistic a perspective that completely ignores potentially counteracting incentive structures, e.g. mutual trust, professional integrity, and a concern for operations and maintenance in the long run. None of these speak in favour of inflating benefits through demand estimates, since there's little point in getting funding for a project that will be too expensive to maintain, or by jeopardizing future funding options due to accusations of unethical behaviour. In fact, since many rail based public transport projects are built in stages, local stakeholders have many incentives to make the initial stage appear as successful as possible after completion. Reports of cost overruns and poor ridership are unlikely to contribute to this.

Even at the national level politicians seem to be aware that the benefits of rail projects are poorly reflected in cost benefit analyses. A parliamentary politician explains that public transport projects are rarely benchmarked against the official requirements because everyone knows that they rarely meet them anyway:

There's always a higher internal rate of return for road projects than for public transport projects. We have this project down south where we expand the motorway and the rail tracks at the same time. Perhaps it would give a higher rate of return to just build the road. But that's a political choice about what to do. Of course the project is not a completely waste of money, but we're below the benchmark threshold... is it 3.5% or something in that range? I can't recall exactly [...] We don't use it to compare road and rail projects, but rather to prioritize between different public transport projects. It's the same with the metro expansion. It's more about whether it's feasible, because in our minds there's no questioning that both the city and the country will benefit from having that. So the transport model is supposed to tell whether the project can support itself financially.

A modeller supports the argument that the overall economic assessments are of little relevance to politicians for these types of projects:

I think it would be interesting to see how low the passenger estimates for a project can get before the politicians go: "Ok, we won't do it". I've never done any cost benefit analysis of the new metro stage. I've helped build the model and seen some calculations from consultancy firms, but I've never seen anything in the papers where politicians go: "2.36... ok, that's too low. We're not building it." So I just figure that the political decision is based on a perception of whether this would be good for the city [...] If a new assessment shows that we get 30% fewer passengers than previously expected, would that result in politicians abandoning the project? I don't think so. I think they want it, and if the internal rate of return is too low they will find other arguments, like they've done with other projects.

The argument here is that the economic performance measures aren't really that important in competition of state funds between conceptually different project types, although it might still be the case among similar project types. However, it is of course still possible that the split in funding between the road and rail sector is influenced by these performance indicators at a more fundamental level. At any rate, the statements calls into question whether cost benefit analyses are at all a suitable tool in providing decision support for policy makers, since the economic benefits in terms of travel time savings appear to have little bearing on the perception of a projects societal value.

Overall, informants identified a number of important aspects in relation to the interpretation of both appraisal results as well as the surrounding contextual framework in which they are embedded. Several informants pointed out that the appraisals cannot be seen in isolation of the processes they are part of, since the results can themselves influence future development and thus the validity of the assumptions upon which they are built. Most informants stressed how policy makers can feel constrained by the rigid structure of existing appraisal methodologies, since they are an important part of the formal decision process but fail to capture many of the benefits that are important to policy makers. This weakens the ability of impact appraisals to function as decision support tools and creates a situation where they often become decision justification tools instead. In line with observations by Henman (2002), it appears that an extensive use of model-based decision support risks hampering the conduct of democratic politics by constructing partisan points of view that are held in undeservedly high esteem compared to their actual validity.

Institutions

The last issue I will present in this section are the institutional structures that surrounds transport project evaluation. Many informants mentioned this as one of the barriers in communication and consequently also in handling uncertainty and inaccuracy in forecasts.

One private consultant explains how the public institutions charged with governing transport planning are often aiming at fulfilling political goals rather than improving transport systems:

Public enterprises are measured based on their effectiveness, and they need to undertake internal improvement processes and stuff like that. They sub-optimize to reach the targets set by the government. I mean, the ministry requested that the rail directorate need to do this and that. They need to fulfil service goals and improvement goals and whatnot... I don't know all the details. But it's pretty clear that decisions are sometimes taken to fulfil a set of goals that are completely unrelated to the substance of a project... that's just how it is. It happens in private enterprises too, but the more the political system spreads down the hierarchy, the more sub-optimising you'll get. Because... I don't know why, but they have this perception that the more they control, the better it gets. And in reality it's the other way around. They get too close. They make too much noise. I would much prefer a set of boundaries and some concrete goals to aim for in the project; the purpose of it. Why do we want this thing? And then they should send some people out to look for answers rather than controlling everything in detail.

The argument here is that people strive to perform based on the performance indicators used to assess their effectiveness. If the performance indicators are not in touch with the actual value production of projects, the organisational structure becomes an obstacle rather than a support in carrying out transport planning of optimal benefit to society. A local politician adds his perspective on how the solutions considered for individual cases are coloured by the professional background of the stakeholders involved:

It would probably be better to use the models for scenario development than for comparison – what happens if we do this or that. It would actually be optimal in relation to scenarios... instead of a model that tells you the truth. There have been some rudimentary efforts to that sort of thing, but it has been toned down. It's not considered interesting. We needed to throw out some hard facts. It's probably also a professional battle among engineers who like to build stuff. Like I say: "Engineers are only concerned with asphalt." Then we have the planners, who bring in some other aspects. No offense to engineers in general by the way, but that's sort of the approach: "We need some fucking roads. We need to get some stuff done." Then the planners have another set of glasses. And then we have people in Parks and Recreation that think about recreational aspects [...] I wouldn't call it a power struggle, but of course there has been a fight about framing the agenda. And there we have to say that Traffic and Roads have had a considerable influence.

In this claim lies an implicit assumption that it is often road and traffic engineers who are employed as transport planners. The claim is then that their educational background predominantly focuses on how to construct new infrastructure as a solution to transport related problems. The claim might apply well in the specific case being addressed, but it is not an assumption I feel comfortable at generalizing from. However, it would certainly prove an interesting inquiry to map the educational background of transport planners in different arenas and how much impact this has on their preferred solutions. At any rate, the citation highlights the tension between different planning sectors (whether based on a real or perceived clash of ideologies) that are influenced by the decisions being made in transport planning for larger infrastructure projects.

The institutional divide was also mentioned in relation to people working with road and rail based transport planning. One civil servant in the rail directorate reflects on his experiences working with the road directorate:

We don't have a lot of collaboration. It's more like we pass on results to each other [...] They do things their way and we do them our way; there's not a lot of partnership in in [...] I think it's because the things that matter in the evaluations are quite different. When you make a road project there are other things that are important than when you make a rail project [...] The rail line doesn't really matter that much to them and their projects. It's kind of a sideshow, where they have to prove once again that rail doesn't matter. That's more or less their approach. I mean, there's ten times the amount of passenger miles for road compared to rail. For them it's just a small uncertainty whether there's a rail line or not. I think that's their mind-set [...] But we had these expansions of a parallel motorway and rail line, where we were surprised about the amount of cars we

would move off the motorway with the rail expansion. First we thought there was an error in the model, but they convinced us that it was probably true. If we take these cars out, then... it's the marginal effect. It's the last cars that cause congestion. It's a balance, and when it tips it doesn't take a lot to make it fall over [...] It can easily have an effect on congestion, even if it's not a lot of cars.

The difference here comes in the form of how a result is interpreted within each organization. The informant claims that the road directorate considers the rail line an negligible uncertainty factor for traffic estimates due to low mode share, while he himself considers even a small increase in mode share as an important goal in reducing congestion. Depending on the magnitude of impact the rail line might reduce the need for expanding the motorway in the first place. However, the road directorate is so accustomed to an exponential increase in road traffic that they more or less consider the status quo mode share as given, and is more concerned with the provision of additional capacity as the system is experiencing mounting congestion. The informant goes on to give an example:

We had a motorway project down south. The road directorate was afraid... they wanted the tracks as far away from the road as possible, right? It can't come too close; otherwise there'll probably be trouble [...] We thought: "Hey, there's only 4 lanes here. We'll just use a bit of this space for something. We'll knock down a wall so we can bring the rail line closer. Then we won't have to expropriate these houses over here." There are some properties where people had already started getting pretty angry. But the road directorate went completely nuts when we suggested that we used some of their road space [...] We were just told to shut up. It was quite rough [...] They already downgraded that section from 6 lanes, but if there were plans to upgrade it again later that would be impossible of course. If we had moved that wall and spent a billion on tracks, well then you obviously can't move that stuff around at a later point.

The point here is more of a turf battle, where each player in the game is unwilling to give up something that could potentially improve the overall transport system if it means an individual loss to them. This is of course not a particularly desirable situation when the agencies are supposed to serve the public at large rather their own interests. The informant goes on to describe different interpretations of more fundamental issues:

We once clashed in a battle unlike anything I've experienced. Some of the most far out stuff I've been involved in. The difference in interests was of course very obvious. From the rail perspective... well, if you want more people to use public transport, the most important thing is to make sure the urban development is condensed into some sort of chunks. That it doesn't create sprawl [...] The road directorate wanted to build beautiful roads in the open landscape instead, where it's cheaper and easier to go with a preferred alignment. We had to make these joint notes for the infrastructure commission with no moderator, but we almost didn't manage. In the end we were correcting formulations when the other part was asleep, and then we passed it upwards before they noticed [...] It was quite important to us how it was mentioned. We wanted the urban sprawl and public transport to be labelled as opposites. They refused that this was the case. They

told us we could just use busses on the motorway. Yeah right! Why don't we use that more often today then? Buses are just semi-poor trains on a road, but the same rules apply as for trains. They need to pass through densely populated areas [...] "What a bunch of nonsense" they said [...] They are road people and we are rail people. And the surrounding environment supports one or the other. That's why – as soon as we discuss the surroundings – then we don't agree on anything.

This citation addresses a more serious problem than mere protection of own interests. The relationship between land use and transport is fundamental in transport project evaluation, since travel demand can be affected through land use planning. Sprawling, low-density cities (e.g. Phoenix) generally have very different modal split and energy consumption for transport compared with compact, high-density cities (e.g. Paris)⁶⁷. A national urban development strategy that is detrimental to the provision of public transport is of course problematic when a goal in many countries is to increase rather than decrease the mode share of public transport. The citation highlights an important issue that is highly relevant to all issues presented in the present thesis, which is a trench war between people working in the road and rail sectors, that might sometimes cause people to interpret evidence in a highly biased manner. It may also create problems of groupthink for projects where no collaboration is forced upon these parties. Several informants brought up the issue of different mentalities between people in each sector. A private consultant working with rail projects explains it as follows:

If you get constructive collaboration going that's great. I have to say that is not always our experience, though [...] Sometimes it's so focused on the budget control that people lose sight of the project's purpose. It's more like a dog fight about some details instead of reaching a good solution. That's regrettable, but those are the commercial terms. It's something else to construct the actual rail line. Most people who work with rail are driven by some kind of – well I don't know what it is – some sort of engagement. A wish about actually... they think railroads are fascinating, fun, exciting, sensible, and all kinds of stuff, right? Most people have some sort of personal involvement in the things they do [...] Passion was the word, that's right.

A consultant that primarily works with roads shares this perception:

One thing is that people working with road transport; they aren't in to cars, they don't have a preference for roads really. They don't have any feelings about it at all. It's just – it's work; it's interesting work – but we're not particular interested in cars. People working in the rail sector, they're usually quite interested in railroads and they really like trains – in a way that you sometimes feel they aren't really neutral to their study object. That is my impression at least.

A researcher agrees and broadens the scope to the political level:

⁶⁷ This should be a fairly uncontroversial claim, but readers might consult Newman and Kenworthy (1989) for an international comparison of fuel consumption for transport and urban density.

It looks like neither the government nor the opposition always formulate the best policies in their own interest. They just come up with random suggestions [...] For instance; the environmental party suggested to build railways in areas, where it's very clear that the use of the railways will be very low. You start wondering: "Why do they suggest this?" [...] I think it would be better for the citizens if the politicians would concentrate more on the aims and the goals for what they think should happen. They tend to suggest precise actions rather than goals and principles [...] They should probably say "We want to make it easier to travel". Instead they say: "We want to build this road", or "We want to build this railway." [...] But they start directly with the actions instead. Maybe that's part of the political system.

Common to all of these descriptions is a central notion of a split between people who work with road and rail projects. Not only are they often separated through the organisational structure, but the mind-set of people working in each sector is also perceived as very different. This extends all the way up to the political level. While passion for your work is not inherently problematic, it creates a risk of overemphasising specific types of policy interventions in situations where they might not be the optimal (or even counterproductive). Examples of pro-road stakeholders advocating motorways as congestion relief in urban areas or pro-rail stakeholders advocating light rail as intercity commute options are sadly not uncommon. Even if these people could agree on whether to build rail or road in a given case, this might overlook important demand management or land use initiatives that could either supplement or replace expensive construction projects. The end result is that society ends up with sub-optimal solutions.

One project manager explains how such sub-optimising strategies are often a distinct feature of public enterprises:

If you build a new cement factory somewhere, well then it's value driven in a degree you couldn't imagine. Building the factory is not what keeps them in business. Starting the production is what generates money. Then you start implementing your benefits, right? That's what keeps them in business. It means that all focus is on benefits. You might lose 500,000 per day until your production is up and running. So the costs? Screw that. They just need to produce stuff. If it costs 2, 5, or 15 million more... fuck that. It's saved in a day. That's a very different way of making decisions, because the entire project organisation and decision hierarchy is aiming for the same purpose [...] We don't have that here. As consultants – let's just be honest – we have our bottom line as the goal. We need to make money of this project, right? But our client, they need to live up to a service contract. And then they do stuff that is measured by their client – in this case the ministry – who have to live up to some policy and produce results that look good in the media. And their client – the government – has long-term strategies. So the purpose points in all directions. That's very different than the industry where everything points in the same direction. So if you lose focus of the value you are creating with a project, then all these different layers of decisions will counteract each other [...] Project errors are never the cause of a dysfunctional project organisation, because there is no such thing as a dysfunctional project organisation. You get exactly the results that you build your

organisation to produce [...] It's the organization. It's the way we don't streamline goals. It's the way we don't follow up on goals. It's the way we are not dynamic about goals. These issues mean we have no clue where we are. We navigate in the dark with pinpoint accuracy.

The point of this argument is that decision-making in planning for public policy involves so many different stakeholders that each bring their own priorities to the table. Stakeholders at different levels have different success criteria for a project, and thus optimise based on whatever success criteria they are presented with. This relates to an earlier argument about the inability of traditional impact appraisal techniques, whether cost benefit analyses or environmental impact assessments, in capturing the true societal value of projects. For all intents and purposes, transport planning cannot function like an industrial enterprise, since the value is an ambiguous term that is not easily measured even if we know the actual outcome prior to implementation.

Overall, informants identified a number of conflicts in the transport planning process that arise as a result of the institutional arrangements of the organisations involved or the professional profile of the people employed there. According to several informants, this can create tunnel vision in the form of groupthink, where stakeholders working within a specific institutional context might be reluctant to consider other approaches to solving a given problem, or other how to define the relevant problem. Furthermore, an organisational structure where rail and road based infrastructure projects are managed by different authorities creates a tension in the evaluation of solutions that favour different modes of transport. The link to forecasting and project appraisal comes from a lack of agreement among the different institutional environments on how the planning of important external factors should be handled, such as policy initiatives for land use and demand management. Since these factors are important in determining the travel demand for different modes of transport, these conflicts become important to the reliability of impact appraisals that are influenced by the choice of planning strategies adopted for these external factors. It seems that many of the conflicts concern the proper treatment of context-related uncertainties (cf. the location dimension described in section 2.3) and how to specify conditions that are external to the models. Uncertainties regarding model specification and calibration are largely technical disputes, which can be reduced by improving the model sophistication and the necessary data for calibration if resources are available. However, the conflicts arising over context and input related uncertainties cannot merely be handled as a technical issue, and risks becoming an institutional battle over ideologies. Given the poor treatment of uncertainties in decision support documents, the specification of context and input can have a large influence on the perceived desirability of different solutions as they are presented to policy makers. Different solutions might be more desirable in other scenarios, but these are filtered out of the possible span of solutions. When the institutions that have traditionally held the power to decide on these matters are unwilling to incorporate new perspectives, there is a risk of non-decision-making (Bachrach and Baratz 1962) regarding the contextual framing. This might cause self-fulfilling prophecies, and it is thus difficult to identify these problems merely from ex-post evaluation of observed forecasting inaccuracy.

7.4 Summary

The purpose of this chapter has been to present the second part of the analysis related to how uncertainty is managed in the policy-making process. The analysis of decision support documents found that uncertainty is generally not communicated very well in the official documentation available to policy makers. The majority of projects analysed had no communication of forecasting uncertainty at all in the available decision support documents. The analysis of questionnaire responses found that demand forecasts are quite influential on how policy makers understand transport related problems, and that result of appraisal is important in negotiations of funding. However, there was little support for strategic misrepresentation being a major explanation of forecasting inaccuracy as a result of these incentive structures. Rather, respondents pointed towards input variables, model structure, data quality, design changes, and coordination of planning decisions as more important explanatory causes. The analysis of interviews supports these findings, but also suggested that political incentives could explain much of the observed forecasting inaccuracy. However, this was not described as strategic misrepresentation in the form of manipulated figures, but rather as selective use of results, disincentives to create awareness of mistakes, and an acceptance of scope-creep no matter what design is initially suggested. In addition, the interview informants also highlighted that many of the most important reasons for building the projects aren't reflected in the appraisals based on demand forecasts, and stakeholders thus come to view these as decision justification tools rather than decision support tools. This becomes especially problematic between different levels of decision-making if the objectives of a project are not streamlined, as different stakeholders will specify assumptions and interpret results in a manner that fits their own definition of the project's objective, but not necessarily those of other stakeholders. As the road and rail authorities are ultimately competing for a limited amount of funds for transport infrastructure projects, this creates a tension that does not foster a particularly corporative environment.

8 Conclusion

Just because something doesn't do what you planned it to do doesn't mean it's useless.

- Thomas Alva Edison

In this chapter I will briefly summarise the findings of the present study in relation to the context of transport project evaluation as it was introduced in chapter 1. I address each research question individually to highlight the most important findings in relation to each of them. The summary is followed by a discussion of the implications of these findings for both planning practice and research, with a point of departure in the review presented in chapter 3. This discussion mainly seeks to address how the validity and reliability of decision support can be improved by better auditing of impacts and communication of uncertainty in light of the findings listed in the summary. It also involves suggestions for future research topics that could improve our understanding of the emergent issues brought forth in the present thesis.

8.1 Findings

I start this concluding chapter by summarising the findings from the analyses presented in chapters 6 and 7. For the sake of clarity I have structured the summary around the research questions introduced in section 1.3, but there will of course be aspects that overlap between these. The presentation of findings related to the first research question is fairly straightforward, since the analytical framework corresponds well with the three sub-questions being addressed sequentially throughout chapter 6. However, the presentation of findings in relation to the second research question requires a bit more manoeuvring through the various parts of chapter 7, as the three sub-questions related to the second research question have been addressed via a triangulation of methods.

8.1.1 Research question 1

The first research question was formulated as follows:

How well do forecasts presented in decision support for transport infrastructure projects match the actual development?

The formulation was split into three sub-questions, which I shall address individually.

1a: What is the extent of forecasting inaccuracy in transport project evaluation?

This question was addressed in detail during the analysis presented in section 6.1. While it was originally intended primarily as a task of preparing a descriptive overview to use as a reference point for more analytical inquiries, it gradually became the dominant unit of analysis both in terms of methodical significance and resource requirements. The reason for this was a considerable number of methodological problems of both theoretical and practical nature that were encountered in

relation to the establishment of a useful measure of forecasting inaccuracy. Previous studies of forecasting inaccuracy have undoubtedly encountered similar problems, but as I did not always agree with the choice of methodology presented in them or was prevented from employing a similar approach due to issues of practicality, I thus decided to devote more resources to the establishment of a proper methodological framework and the necessary data items to pursue it that initially planned. That being said, it was also of key importance to arrive at results comparable with those of previous studies. Conversely, the analysis of the reminding research questions were devoted fewer resources than initially planned, and ambitions were necessarily scaled down for these.

For road projects, the analysis showed that forecasting inaccuracy for demand was biased towards an underestimation of traffic volumes, albeit with a large degree of imprecision that makes it difficult to correct individual projects for this bias. On average, actual traffic volumes were 10% above the forecasted figures. Furthermore, the median value was remarkably close to zero, indicating that the observed bias is caused by several projects with quite considerable underestimation of traffic. This means that road projects are, *ceteris paribus*, equally likely to experience demand levels below and above those estimated in traffic forecasts. The results are very identical to those presented in previous studies of forecasting inaccuracy for road projects, with the exception of a study of British projects that were completed during the oil crises of the 1970s. Forecasting inaccuracy for cost was biased towards underestimation, with actual costs being 16% above estimates on average. This trend is in line with results from previous studies, and the magnitude of bias falls somewhere in the middle of these prior results.

For rail projects, the analysis showed that forecasting inaccuracy for demand was biased towards an overestimation of passenger volumes, although this is also associated with a large degree of imprecision. On average, actual passenger volumes were 20% below the forecasted figures. Furthermore, the media value was also around 20% below the forecasted figures, indicating that the observed bias is caused by a general tendency of demand overestimation in forecasts. This means that rail projects are, *ceteris paribus*, more likely to experience demand levels below those estimated in traffic forecasts. The results are identical to those presented in previous studies of forecasting inaccuracy for rail projects, although the magnitude of observed bias is considerably less in the present study. Forecasting inaccuracy for cost was biased towards underestimation, with actual costs being 14% above estimates on average. This trend is in line with results from previous studies, but the magnitude of bias is somewhat lower than in these prior results.

For zero-alternatives, the analysis showed that forecasting accuracy for demand was biased towards an overestimation of traffic volumes, and with a considerably lower imprecision than observed for the analysis of completed projects. On average, actual traffic volumes for the zero-alternatives were 7% below the forecasted figures. Furthermore, the median value was roughly 6% below the forecasted figures, indicating that the observed bias is caused by a general tendency of demand overestimation in forecasts. This means that zero-alternatives are, *ceteris paribus*, more likely to experience demand levels that are below those estimated in traffic forecasts. To the best of my knowledge, this is the first empirical study on forecasting inaccuracy of zero-alternatives, and there are thus no previous results against which to compare.

1b: Is it possible to identify characteristics of projects that are prone to forecasting inaccuracy?

This question was addressed during the analysis presented in section 6.2. Due to the difficulty in obtaining the necessary data, it has not been possible to perform a detailed regression analysis to identify the most important indicators of forecasting inaccuracy. Instead, the analysis focused on investigating some of the most likely indicators among the available data items for the projects in the sample. This was supported by a more qualitative analysis of groups of projects with various degrees of forecasting inaccuracy. It should be noted that the analysis of these characteristics are confined to demand forecasts.

There appeared to be no clear improvement in forecasting accuracy over time for road projects. However, there was a noticeable reduction of bias for rail projects, even when accounting for the large imprecision observed for both project types. This indicates that a mounting focus on the problem of overestimated forecasts could possibly have had an influence on practice, although the trends are much too vague to conclude anything but crude indications for this to be the case. There also appeared to be no strong relationship between forecasting inaccuracy and time horizon of forecasts, although the most extreme levels of inaccuracy were found among projects with the longest timeframes. This indicates that transport project appraisal seeking to evaluate impacts more than 8 years into the future are associated with considerable risk, and those seeking to evaluate impacts more than 15 years into the future are likely rough indicators at best.

When splitting the samples for road and rail projects into different project types it was possible to identify a set of new patterns. There was a clear tendency for upgrades of existing links to be associated with considerably less forecasting inaccuracy compared to construction of new links, both in terms of bias and imprecision. This was true for both road and rail projects and indicates that forecasting of future demand is relatively unproblematic when no new links are added to the transport network. In addition, for road projects there were considerably higher levels of forecasting inaccuracy for bypasses and fixed links compared to motorways, both in terms of bias and imprecision. This indicates that these project types are more disruptive to transport patterns than motorways. Alternatively, it could be an indication that limited resources are devoted to forecasting demand for these project types, which is likely to be a more plausible explanation for bypasses than for fixed links.

In relation to the hypothesis of strategic misrepresentation, little to no support was found for this in the relationship between forecasting inaccuracy for demand and cost estimates, as these appeared to be largely uncorrelated. The exception was for new motorway links and fixed links, where there seemed to be a weak positive correlation between the two types of forecasting inaccuracy. However, the complexity in using travel demand as a proxy for benefits makes it difficult to estimate whether this is also the expected trend we would observe from strategically misrepresented forecasts, just as other plausible explanations would account for the observed correlation.

In the qualitative analysis it was further emphasised that project type might be a reasonable proxy for the inaccuracy level, but a few other patterns also emerged. The road projects that experienced the most significant underestimations of demand were either fixed links or crucial motorway links that completed a larger network. This supports the previous finding that the most extreme deviations are indeed found among the projects that cause the largest disruptions to a network. Projects with more moderate levels of underestimation were typically new motorways or bypasses,

where induced traffic had been given little attention or the local growth rate had far exceeded national levels. Among the group of road projects with the best accuracy there was a clear dominance of upgrades to existing links. The projects with an observed overestimation of traffic had a lot of bypasses, where traffic levels on the overall network was considerably more accurate than on the specific project link. This indicates that overestimation of traffic for road projects might be caused by poor distribution models rather than an overestimation of actual traffic demand.

The qualitative analysis of rail projects found that the projects attracting more passengers than expected were typically located in denser urban areas and completed in the latter half of the investigation span (1970-2010). This indicates that ridership forecasts for urban rail projects are improving and/or that there is a change of travel patterns towards a higher mode share for public transport in the case countries. The projects with the most accurate forecasts are almost exclusively upgrades to existing systems, further supporting the earlier finding that inaccuracy is mainly problematic in relation to larger disruptions to the transport system. Projects that experience considerable ridership shortfalls compared to expectations are typically inter-city links or airport express services, where there has been a marked optimism in the ability of the new service to outcompete existing road-based alternatives. This indicates that problems of overestimated rail forecasts might primarily be associated with failed expectations for rail services to replace bus or car commuting on non-congested networks.

1c: What is the influence of forecasting inaccuracy on subsequent impact assessments?

This question was addressed during the analysis presented in section 6.1.3. A comprehensive reappraisal of the projects used to measure forecasting inaccuracy would be much too resource demanding, and since considerable resources had already been spent on question 1a, I decided to do a brief analysis of two critical cases; one rail and one road.

Ringbanen in Copenhagen was selected as the rail case due to its considerable cost overrun and ridership shortfall. This places it in the same category as the projects that have motivated a massive criticism of forecasts for rail projects grossly misrepresenting the likely benefits. The analysis was somewhat shallow, being mainly based on reflections over the quality of decision support and the perceived success of the project as it is communicated in press releases and the media. It was clear from the decision support that optimistic ridership figures had been selected from a range of possible outcomes, and that the forecast used in the economic appraisal was artificially high. However, the project is still perceived as quite successful, and most of the additional costs were due to politically approved add-ons rather than the initial project design being more expensive than expected. It has also been difficult to assess whether the lack of ridership on Ringbanen is simply due to some of the traffic flow being served by existing rail links in the same corridor.

Nordhavnsvej in Copenhagen was selected as the road case due to its placement in an already congested part of the Danish capital region. The case was meant to illustrate how additional traffic compared to expected figures is not necessarily associated with increased benefit in the form of travel time savings. By comparing the results of cost benefits analyses based on two different transport models, it was shown that a mere 5% additional traffic would result in a 40% less benefit in the form of travel time savings. In addition to illustrating the complexity of using forecasting

inaccuracy of demand as a proxy for the accuracy of benefits estimates, the case also illustrated the problem of adequately addressing induced demand effects in transport project evaluation.

8.1.2 Research question 2

The second research question was formulated as follows:

How is uncertainty managed in the policy-making process?

The formulation was also split into three sub-questions, which I shall address individually.

2a: How are uncertainties communicated between stakeholders?

This question was addressed throughout all sections in chapter 7, but with focus on different aspects of communication in each section. The aspect being focused on in each of these sections was chosen with regard to how suitable the respective source of data presented to the specific section was at addressing a particular type of communication.

Section 7.1 focused on the communication of forecasting related uncertainties as they are presented in the official decision support documents made available to policy makers. The findings showed that uncertainty in relation to demand forecasts is generally not communicated at all, and the vast majority of available documents simply present expected traffic volumes as a given figure. In relation to cost estimates there was a higher focus on presenting quantified risk assessments for the results, although the majority of projects still had no communication of uncertainty.

Section 7.2 focused on how stakeholders in general perceived the communication of uncertainty and the role of forecasts as decision support. The findings showed that stakeholders will generally accept a moderate degree of inaccuracy in forecasts, but that they feel uncertainties to be poorly communicated in forecasts. However, politicians were generally less concerned with the lack of documentation than other groups of respondents. In addition, the findings showed that stakeholders consider transport models a necessity for understanding the complex systems that are affected by transport infrastructure planning, that the forecasts produced are quite influential on decision-making, and that important political priorities are often ignored in appraisal.

Section 7.3 focused on the emerging concepts from the interviews with key stakeholders. One of the most frequently addressed concepts by informants was the communication of uncertainty, interpretation of results, and the institutional framework. Informants pointed out that communication of uncertainty is generally given sparse attention by modellers, who are mainly committed to establishing a model that produces reliable results. In addition, policy makers were considered to be poor at grasping uncertainties when presented with them, which results in a lack of request for detailed uncertainty communication. Finally, consultants expressed reluctance in engaging in open discussions about uncertainties in fear that it might compromise their professional integrity as independent advisors when their contributions are misinterpreted as partisan input.

2b: What are the typical explanations offered for inaccurate traffic forecasts?

This question was addressed in sections 7.2 and 7.3 by focusing on different sources of evidence for the explanations being offered by stakeholders.

Section 7.2 focused on the questionnaire responses to a set of predetermined explanatory categories, which were presented individually and ranked via a Likert-scale response format. The findings show that stakeholders generally perceive strategic misrepresentation in the form of manipulated figures to have a fairly low explanatory power. Instead, respondents point to unexpected development for key input variables as the most important explanation, as well as design changes after approval, poor quality of data for input and calibration, and neglect of induced demand effects.

Section 7.3 focused on the emerging concepts from the interviews with key stakeholders. The three most frequently addressed concepts by informants were neglected effects, unexpected development, and political incentives. The neglected effects mainly related to known issues that the models are poor at reflecting, such as induced demand as well as public and non-motorized transport modes. The resource-intensity of developing and maintaining models was mentioned as an important reason for this neglect. The unexpected development mainly related to poor data quality for inputs and parameter estimations, or to changing planning frameworks due to new political commitments that has not been included in the project appraisal. The political incentives mainly related to selective use of results, failure to disclose mistakes in the appraisal, and acknowledgement of scope-creep regardless of initial project design.

2c: Which obstacles do stakeholders identify in the use of traffic forecasts?

This question was mainly addressed in section 7.3 and based on the emerging concepts from the interviews with key stakeholders. However, the analysis of this sub-question has probably suffered the most from the additional resources devoted to research question 1a, and I therefore advise critical scrutiny in the interpretation of these findings. The two most frequently addressed concepts in this regard were interpretation and institutions.

Regarding interpretation, informants pointed out that appraisal results can themselves be used to change the likely development depending on the interpretation of them. The obstacle aspect becomes relevant when certain stakeholders interpret the results as prophecies rather than forecasts, and blindly cater for the predicted demand instead of considering if the demand itself be affected. Another critical aspect that several informants mentioned was the lack of correspondence between important political priorities and the benefit categories included in impact appraisal. Since many policy makers do not see their values reflected in the appraisal framework, the results of impact appraisals carry little weight in shaping their decisions.

Regarding institutions, informants pointed out that a set of unhappy dualism exists in the institutional framework of transport infrastructure planning, where rail and road proponents often compete for the largest market share rather than collaborate to achieve the most effective solutions. This was mentioned both in relation to modellers, planners, researchers, and politicians alike, with the result being a sub-optimisation of individual projects due to conflicts of interest.

8.2 Discussion

As described in chapter 1, the overall topic of the present thesis was the validity and reliability of ex-ante appraisals for transport infrastructure projects as they are used in decision-making processes.

More specifically, the analysis was aimed at the inaccuracy of forecasts in these appraisals and the management of uncertainty by key stakeholders during the process. As has been summarised in section 8.1, forecasts display a high degree of inaccuracy and uncertainty is often poorly communicated among stakeholders. I shall spend the remainder of the concluding chapter on discussing the implications of these findings as well as future research topics that could improve our understanding of these issues. These reflections will also address the findings of previous studies presented in chapter 3 as well as the experiences in relation to data access that were described in chapter 5. The discussion is built around a set of highlights that can be concluded based on the findings listed in section 8.1.

8.2.1 Reinterpreting forecasting inaccuracy

As discussed in chapter 3, ex-post evaluations of rail projects have typically found forecasts of demand and costs to be overly optimistic. The review included in the present thesis mainly dealt with the few high profile studies based on larger samples, but a similar trend can be observed when aggregating findings from individual case studies (Brinkman 2003). The results in the present study indicate some support for a trend of optimistic forecasts, but the magnitude of the observed bias is much smaller than in previous studies. This might be due to an improvement in the forecasting practice, since results indicate that newer projects are on average less biased than older projects. However, it might also be the case that previous studies of forecasting bias in rail projects have not been very representative for typical rail projects. All the previous studies reviewed in chapter 3 displayed considerable availability bias in their sampling, focusing on projects where local expertise was limited due to little or no previous experience with these types of projects in the case country, or where data was available due to the very fact that the forecasts had proven to be biased. These observations are in themselves interesting, but not very useful in assessing the bias associated with more standardized rail projects. It is thus problematic that these studies have been used to promote a general scepticism towards the feasibility of rail-based public transport systems with little or no regard to the type of projects that these results refer to⁶⁸. Conversely, the present study has used a more randomised sampling approach, and the results are thus more representative of a typical rail project in the four case countries.

The demand forecasts presented in the decision support for rail projects are typically also based on favourable assumptions regarding future land use planning. The urban structure is quite important in determining the feasibility of public transport modes, but often the development fails to meet the expectations specified in ex-ante appraisal. This has especially been the case in the US, where rail investment has typically been followed by highway orientated development (TCRP 2004). This might explain why some of the previous studies have found a higher forecasting bias than the present

⁶⁸ There are plenty of examples of uncritical reference to these results. A recent example is Cox's (2012) reassessment of the XpressWest project for the Reason Foundation, which is conducted by adjusting ridership figures downwards based on an 'International Average Error Forecast' factor without any discussion of why this adjustment might be relevant for the XpressWest project. This adds no useful information and only reduces the ability of the appraisal to function as decision support. I assume the adjustment factor is based on the results from Flyvbjerg et al. (2005), as the figures are identical. However, the approach used by Cox is far too simple, and not in line with Flyvbjerg's suggested principles of reference class forecasting. A rail line through a flat area with no grade separation is hardly in the same reference class as metro systems built in developing countries during the 1980s. Having 'rail' as a reference class is thus not of much use when it is applied so carelessly.

study, since many of the projects included in these have been completed in the US during a period of highway orientated development, while the forecasts have assumed transit orientated development. Johnston (2006) even argues that it is necessary to allow highways to become congested if planners are to utilize the full potential of transit options. However, as several stakeholders pointed out, politicians are often unwilling to commit to such drastic changes, since this would impose a perceived penalty on car users, who often constitute the majority of voters. The ridership shortfalls could therefore also be a reflection of a political reluctance to facilitate the necessary conditions for rail projects to perform as described in decision support. It might be theoretically possible to reach the expected ridership figures, but it might be entirely impossible to find political support for the land use planning necessary to facilitate them. This is why the necessary conditions for the expected ridership need to be clearly specified in decision support, so that policy makers may evaluate how plausible these assumptions are. If the tendency for demand forecasts to be overestimated in impact appraisals for rail projects is indeed due to the specification of optimistic assumptions by rail enthusiasts (whether intentional or not), increased transparency regarding the underlying assumptions is a necessary, albeit likely insufficient, first measure towards remedying this problem.

Concerning road projects, the present study found that demand forecasts are generally underestimated, in line with most previous studies. However, the bias is not as large as for rail projects, and since more traffic tends to materialise than expected, this has sparked further rail criticism, claiming that funding for roads ought to be increased since the demand is rising faster than expected. However, these arguments ignore the conditional land use planning that has often favoured road orientated development in the periods where these rail projects have opened. Furthermore, they ignore the fact that additional traffic is not in itself a benefit, since it often incurs congestion and adverse environmental effects. As has been shown in the present study, congestion incurred by additional traffic can also reduce the economic benefits quite drastically. This does not mean that no benefit is derived from a new project, but it illustrates the problem of using demand as a proxy for benefits. Many of the road projects with overestimated demand forecasts are also bypass constructions, where relatively simple models have often been used to assess the distribution of traffic. In the present study there were several examples of demand for the total network being underestimated in these projects, which indicates that the true forecasting bias in road projects might be larger than reported. It also indicates that the congestion relief and noise reduction from bypass construction is likely overestimated, since more traffic than expected will continue to pass through the urban settlements the bypass was meant to relieve.

The present study also found that forecasts for zero-alternatives are systematically overestimated, which is important for the economic appraisal of projects since this forms base value of comparison for all alternatives being assessed. By overestimating demand in the zero-alternative the appraisal assumes higher levels of congestion than would actually be the case, and a capacity expansion thus appear as a more feasible investment than would otherwise be the case. The result is an inflation of the internal rate of return in assessments of road projects, which make alternative approaches to congestion relief appear less desirable from a socioeconomic perspective. This might be considered an example of a reification fallacy, where failure to accommodate rising demand is believed to result in increasingly rising travel time for commuters due to congestion. In reality, some maximum travel time that new users are willing to accept exists, and traffic will thus not continue to rise exponentially

as congestion worsens. However, the forecast for the zero-alternative often disregards this, and the cause of the fallacy comes from trusting the model results despite disagreeing with the assumptions.

This fallacy is easy to spot when we consider the mechanisms involved, but in appraisal this assumption is often hidden and thus rarely contested by stakeholders. An example is the long term forecast for road traffic prepared by the Danish infrastructure commission, which made a crude projection of future travel demand under the assumption that there was no capacity restraints on the road network (IK 2008). This demand was then distributed on the existing network to identify where the most critical bottlenecks would appear. This approach is useful only insofar that a principal decision has already been made to solve all congestion problems via road capacity expansions, and traffic should be allowed to grow unconstrained. However, as illustrated by Næss et al. (2012), such an approach would quickly lead to the need for 10 or more lanes in each direction on the motorways in larger urban areas.

A true zero-alternative would not actually experience the level of growth portrayed in this projection, as future growth in population and employment will start seeking less congested transport axes or transport modes, e.g. by locating near transit hubs or traveling by bike. However, this is typically completely ignored by modellers, and the results from the commission are now used as a baseline for the expected baseline growth in zero-alternatives for new road infrastructure in Denmark, and the Danish road directorate counters criticism of this assumption simply by referring to the commission's report (VD 2012). A slippery slope fallacy is thus institutionalised in appraisal practice and reinforced via circular reasoning. The perceived accuracy of demand forecasts in road projects compared to rail projects thus might be due to self-fulfilling prophecies⁶⁹ or accidental truths⁷⁰ rather than superior quality of decision support⁷¹. This is not surprising, as travel demand forecasting is *"orientated almost exclusively toward analysis of long-term, capital-intensive expansion of the transportation system, primarily in the form of highways"* (Pas 1995, 55). It is, however, problematic if the political priority is not simply to 'predict and provide' road capacity, but rather to 'predict and prevent' exponential growth in traffic as Owens (1995) puts it. As was the case for road projects, increased transparency regarding the underlying assumptions is a necessary step if this problem is to be remedied.

When discussing forecasting inaccuracy, bias in the form of the mean inaccuracy value often gets the most attention, while the imprecision in the form of the spread is overlooked. However, the spread is equally important in relation to the validity and reliability of impact appraisal. Even in the absence of bias the observed imprecision is large enough to have a considerable impact on the results of appraisal. When raising this issue with planners, a typical response is that the problem is less severe, since appraisals are mainly used to compare the relative value of projects against each other. The problem with this argument is that it assumes all projects to be equally affected by this imprecision, which is not always the case. Obvious examples are comparisons of conceptually different projects,

⁶⁹ See Merton (1968) for an introduction to the concept of self-fulfilling prophecies and Næss (2011) for an illustrative case study in relation to transport infrastructure planning.

⁷⁰ See Reed (2000) for a discussion of accidental acquisition of knowledge.

⁷¹ As Russell (1967) famously illustrated with reference to Ptolemaic astronomy, predictive accuracy is not in itself a measure of the explanatory power of a theory.

comparisons of projects in different geographical areas, and comparisons of new schemes with a zero-alternative. If both costs and benefits are likely to deviate greatly from forecasts, there is little use of ranking projects based on their internal rates of return or other economic performance indicators, unless the differences are quite substantial.

8.2.2 Systematic evaluations are necessary

As discussed in chapter 5, reliable data sources for the analysis of forecasting inaccuracy has been difficult to obtain. This is a problem shared by all previous studies listed in chapter 3, and it makes it challenging to provide detailed analysis of the extent of inaccuracy as well as the likely causes. From a research perspective, the problem mainly relates to the ability to undertake studies of interest to planning theory and practice. Retrieval of the necessary data is simply too resource demanding for most scholars to bother with original data collection for large samples, which is also why the vast majority of ex-post evaluations are based on individual cases or small sample. As discussed in chapter 3, this creates an unfortunate availability bias when these are aggregated into larger samples, leading to a false sense of statistical significance regarding results. Such aggregation also relies on similar data items being reported in each study, which is far from the case. This result in datasets that lack the necessary information required to control for a variety of possible determinants of inaccuracy.

From a planning perspective, the main problem is that an important learning potential is lost, since this data can provide valuable information about how systems react to the disruption caused by new infrastructure projects, as well as what went wrong in the appraisal process if expectations failed to be met. However, when no systematic ex-post evaluations are undertaken, this learning process is severely limited, and often the explanations being offered in hindsight are based on guesstimates rather than analytic enquiry. In this regard it is no surprise that forecasting accuracy hasn't improved over the last decades. Some scholars, who stress the importance of strategic misrepresentation as an explanation for forecasting bias, claim that planners should have learned from past mistakes if bias is truly innocent (e.g. Cantarelli et al. 2012; Flyvbjerg 2007), but such arguments ignore the lack of a proper evaluation systems to enable a learning process. Several scholars have stressed that stable environments and rapid, unanimous feedback in the form of systematic evaluation is necessary in order to facilitate a learning process from failures (Hogarth and Makridakis 1981; Slovic 1972). The experiences from the transport planning sector in all four case countries indicate that none of them have been able to meet these requirements in the periods covered by studies of forecasting inaccuracy⁷².

From a broader societal perspective, the main problem is that there is no systematic performance evaluation of what benefits are obtained from transport infrastructure projects. To put it bluntly, we don't know a whole lot about what we get for our taxpayer money. The present thesis has mainly been addressing demand forecasts as a benefit estimate, but the forecasting accuracy related to travel demand is just a proxy for the true societal benefits (or costs) derived from transport infrastructure. Without a systematic evaluation system, we don't really know if travel time

⁷² The POPE reports are an exception, as the assessments are reasonable standardized and cover all schemes above £5 million. However, this is a more recent programme that has had no influence on the previous studies referred to in chapter 3. Furthermore, the evaluation is limited to road construction rather than being a multimodal transport evaluation programme.

decreased, air quality increased, or traffic safety improved as a result of a project. Large amounts of resources are spent on constructing and calibrating models to prepare appraisal results. In contrast, relatively few resources are spent on figuring out whether the results of the model were correct. The general mentality seems to be that monitoring and evaluation are optional activities, and that it is more important to move on to the next assignment. After all, the project has been built, the budget has been spent (probably more than so), and the evaluation won't change any of that. The result is that stage 7 in the policy cycle (Figure 4, page 14) is usually skipped, and as a result the quality of future decision support and peer review is unable to benefit from the learning that should have taken place here.

One particular recent area of planning where the need for systematic ex-post evaluations becomes crucial is the use of quantified risk assessment procedures, e.g. by reference class forecasting or similar approaches. The basic premise is to adjust the risk assessment of projects based on past experiences with forecasts for similar projects. Such an approach could greatly reduce optimism bias, as it allows planners to adopt the 'outside view' (Kahneman and Tversky 1979). However, such an approach has a number of shortcomings. First, with no systematic ex-post evaluation, no reliable reference data exists for such an outside view. The present study, as well as the results presented in chapter 3, is perhaps state-art-of-art, but in lack of more detailed reference data it would only be possible to produce very crude risk analysis with a reference class approach. Second, the approach is based solely on aggregate data⁷³, and would not provide much new knowledge about what the cause of uncertainty is. This is not the intention of the approach of course, but this causes a risk of overlooking qualitative aspects that might be more important. For example, reference class forecasting was used by the consultancy firm Ove Arup and Partners to conduct a risk assessment of the Edinburgh trams project (Arup 2004), which adjusted the cost estimate from £255 million to \$400 million. However, in 2011 the capital expenses were already up to £461 million, with estimates for total costs passing £1 billion (CEC 2011). In the context of such massive cost overruns, the average bias is largely irrelevant. It would be much more important to figure out how to identify the risk of running into such 'black swans' (Taleb 2010). Third, a reference class forecasting approach naturally encounters a well-known reference class problem; how to identify a proper reference class for a given item. Reichenbach (1949, 374) formulates the problem as follows:

If we are asked to find the probability holding for an individual future event, we must first incorporate the event into a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result.

In relation to transport project evaluation we are faced with the problem of how to establish reference classes that are suitable for a new project. It seems problematic to consider all road projects or all rail projects identical, as the different subgroups of project types have different probability distributions for inaccuracy (cf. chapter 6). Furthermore, there might be important circumstances that a given project unsuited for comparison with past cases. As an example, the competition between different modes is often very different in Denmark than the other case

⁷³ This could probably be remedied by applying it to individual risk items, but this would also increase the data requirements exponentially. Given the data availability this is also unfeasible at the moment.

countries in the present study, since the mode share for bikes is much higher. Modal shift experiences from light rail projects in Norway might therefore not prove very useful for evaluating potential modal shift for light rail projects in Denmark.

8.2.3 Precision versus pedagogy

An interesting observation was made while conducting interviews for the analysis presented in chapter 7. Almost every informant indicated that policy makers consider the value of projects in a much broader framework than what is included in cost benefit analyses. Meanwhile, most modellers are concerned with improving the sophistication of transport models to account for even more detailed aspects of travel demand. The immediate problem with this is of course that increasingly sophisticated models require increasingly detailed input and calibration data. A more sophisticated model might have fewer specification errors, but it might increase the likelihood of measurement errors. Van Wee (2011) notes that many of the sophistications claimed to be possible by modellers are only convincing under assumptions of unlimited resources and disregard to measurement errors. Similarly, Kay (2011) argues that there is a general tendency in economic disciplines to build sophisticated models that require information that we cannot provide. As data quality is already problematic for existing models, it is doubtful whether increased model sophistication is going to improve forecasting accuracy unless additional resources are also devoted to upgrading such models regularly.

Even if data was readily available to calibrate such models, external variables would remain a large uncertainty factor. As outlined in chapter 2 there are some inherent limitations to the predictability of social systems, and many of these stem from mechanisms that are either impractical or impossible to include in model-based decision support. The resources spent on increased model sophistication thus might improve accuracy, but whether the gains would be worth the cost is less certain since there are many other sources of uncertainty than model specification. Ramsey (2005) argues that our concern with model sophistication causes a situation where *“we are focusing on the mice under the furniture while ignoring the elephant in the middle of the living room”*. If we could somehow predict how people would behave in any given scenario, we would still be left with the task of specifying the most likely scenario; a task that might involve a lot more uncertainty than using a less sophisticated model.

The typical approach to this problem is to assume an ergodic social system, where observations of the past are the best indicators of the most likely development in the future. Such an approach is insensitive to fundamental shifts in behaviour as well as the option for policy makers to actively affect behaviour through policy initiatives. A contemporary policy agenda in many countries is specifically to invoke behavioural changes for transport activities, but increased model sophistication could risk enforcing problems of assumption drag. Since this sophistication likely results in models that are also increasingly cumbersome to adjust in face of changing trends in society, it means that trend extrapolation of outdated transport patterns risk being continued longer than would otherwise be the case. In addition, increased sophistication would most likely increase the black-boxing effect of model-based decision support, causing policy makers to feel even more remote to the appraisal being prepared for them. Since policy makers already seem largely concerned with things that are not even included in the models, the focus on model sophistication as a remedy to forecasting

inaccuracy seems to be a costly approach that is only able to add limited value to decision support. Simpler, theory-informed models might be a better alternative to focus on, with the goal of having policy makers realise important system mechanisms and understand how different approaches can be used to reach different goals. These should involve multiple scenarios for the development of external variables and include more comprehensive sensitivity analyses, rather than pretending to calculate the exact impacts of a project.

This approach would be a step towards acknowledging that we are unlikely to produce very accurate forecasts, and by focusing on simpler models it will hopefully be easier for both experts and laymen to submit appraisal results to critical scrutiny. However, such an approach does not square well with the function of model-based decision support in its current form. Cost benefit analyses require inputs to be quantified with extreme accuracy, and this is probably the reason why continued model-sophistication is considered a necessary advance. Still, it is questionable if these forecasts are able to predict whether a new road will result in commute times that are on average 2.15 or 3.12 minutes faster, but such a difference could mean a difference in travel time savings of several hundred millions. Meanwhile, many other beneficial aspects of these projects are ignored in the appraisal; we simply measure what is measurable, and leave out the rest.

It is understandable that policy makers request a structured approach to rank projects against each other, and the argument of using performance indicators only as relative figures holds some merit in cases where a decision on a specific solution has already been chosen. The solution could then be to use the simpler models in the scoping phases to compare conceptually different solutions. When a decision has been made to go ahead with a specific solution, more sophisticated modelling and cost benefit analyses could be introduced. At earlier stages the complexity introduces so many additional uncertainties that they do not add much value to the decision support. When the cost benefit analyses is used mainly as a tool to justify decisions rather than support the decision process, policy makers continuously seek information that might improve their confidence but which adds little accuracy to the appraisal (Oskamp 1965; Slovic 1972).

A simpler approach would hopefully do away with some of illogical assumptions that persist in transport planning. The assumption of continued baseline growth discussed in subsection 8.2.1 might have been useful when the primary goal of transport planning was to construct a nationwide system of highways and motorways in the post-war era. However, it is a counterproductive assumption for many of the contemporary transport challenges, such as how to provide effective mobility options in congested urban areas. The assumption builds on the premise that planning can or should not interfere with levels of demand, but merely with levels of supply. In other words, the current appraisal system favours the status quo approach and disfavours projects that seek to bring about a behavioural change (Ramjerdi 2011), even when this is in fact a stated policy goal in many countries. This is a very ineffective way of approaching problems related to congestion, social exclusion, or CO₂ emissions, but since this assumption is buried in the models, stakeholders are consistently deceived to believe that increasing capacity is the most appropriate way to deal with these issues. Simpler models would hopefully have more difficulties incorporating such counterproductive assumptions. This would hopefully release some of the many resources spent on detailed modelling, that could then be spent on other areas of uncertainty management. As was discussed in chapter 7, the

uncertainties related to context and input are often largely ignored, while they are also some of the areas that policy makers have an ability to influence through their decisions in other areas (e.g. land use planning or pricing schemes).

Are forecasts then fact or fiction? Unsurprisingly, the answer is a bit of both. Plenty of facts are thrown around in the preparation of demand and cost forecast, but at the end of the day the validity and reliability of these rely mainly on the tales told by stakeholders. They might tell tales of urban development that never materialize, or they might speak of high transit frequency that ends up with twice the headway. The results are thus often as fictitious as the tales they are based upon, but during the process the stories are translated into a language that few stakeholders understand. A set of final recommendations for how to improve the validity and reliability of forecasts must therefore be to focus on simpler, theory-informed models to describe complex system behaviour, transparent documentation of assumptions and specifications, and systematic ex-post evaluation of the expected versus actual outcome of projects. Especially the uncertainty dimensions that lie outside the models need an increasing amount of attention, since these are often more important for the validity of results than those inside the models, but largely ignored in model-based decision support documentation.

9 Bibliography

- Ackerman, F., and L. Heinzerling. 2004. *Priceless. On knowing the price of everything and the value of nothing*. New York: The New Press.
- Ackoff, R.L. 1999. *Re-Creating the Corporation*. New York: Oxford University Press.
- American Association of State Highway Officials. 1957. *A policy on arterial highways in urban areas*. Washington: American Association of State Highway Officials.
- APTA. 1991. *Fare elasticity and its application to forecasting transit demand*. Washington, DC: American Public Transit Association.
- Armstrong, J.S. 2007. 'Significance tests harm progress in forecasting'. *International Journal of Forecasting* 23(2): 321–327.
- Arup. 2004. *Edinburgh Tram Line 2. Review of Business Case*. West Lothian.
- ATKINS. 2009. *A500 Stoke*. Birmingham: Highways Agency. One Year After Study.
- . 2005. *A6 Rushden Higam Ferres Bypass*. Highways Agency. One Year After Study.
- Bachrach, P., and M.S Baratz. 1962. 'Two faces of power'. *The American Political Science Review* 56(4): 947–952.
- Bain, R. 2009. 'Error and optimism bias in toll road traffic forecasts'. *Transportation* 36(5): 469–482.
- . 2010. 'S&P Project Data', personal email correspondence.
- Banister, D. 2008. 'The sustainable mobility paradigm'. *Transport Policy* 15: 73–80.
- Banister, D., and M. Thurstain-Goodwin. 2011. 'Quantification of the non-transport benefits resulting from rail investment'. *Journal of Transport Geography* 19.
- Bates, J. 2000. 'History of demand modelling'. In *Handbook of Transport Modelling*, Oxford: ELSEVIER, p. 11–33.
- BD. 2004. *Flintholm Station er åbnet*. Danish Rail Directorate (Banedanmark). Press Release. <http://www.bane.dk/visNyhed.asp?artikelID=2109> (Accessed March 29, 2012).
- Berlingske. 2008. 'Ekspert: Ny Øresundstunnel nødvendig'. *Berlingske*. <http://www.b.dk/koebenhavn/ekspert-ny-oeresundstunnel-noedvendig> (Accessed March 26, 2012).
- Beukers, E., L. Bertolini, and M.T. Brömmelstroet. 2012. 'Why Cost Benefit Analysis is perceived as a problematic tool for assessment of transport plans: A process perspective'. *Transport Research Part A* 46: 68–78.

- Brinkman, A.P. 2003. 'The Ethical Challenges and Professional Responses of Travel Demand Forecasters'. Ph.D. Thesis. University of California.
- Bristow, A., and J. Nellthorp. 2000. 'Transport project appraisal in the European Union'. *Transport Policy* 7(1): 51–60.
- BS. 2000. *Ringbanen Miljøredegørelse*. Copenhagen: Danish Rail Directorate (Banestyrelsen). Environmental Impact Assessment.
- Button, K.J. et al. 2010. 'The Accuracy of Transit System Ridership Forecasts and Capital Cost Estimates'. *International Journal of Transport Economics* 37(2): 155–168.
- BV. 2006. *Effektuppföljning av Mälärbanan*. Swedish Transport Administration (Banverket). Ex-post Appraisal.
- Cantarelli, C.C. et al. 2012. 'Characteristics of cost overruns for Dutch transport infrastructure projects and the importance of the decision to build and project phases'. *Transport Policy* 22: 49–56.
- CEC. 2011. *Edinburgh Tram Project*. The City of Edinburgh Council.
- Clark, B. 1998. *Political-economy: A comparative approach*. Westport, CT: Praeger.
- Collier, A. 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. Verso.
- COWI. 2008. *Hvor blev de af? Passagerudviklingen i hovedstadsområdet 2002-2007*. Lyngby.
- Cox, W. 2012. *The XpressWest High-Speed Rail Line from Victorville to Las Vegas: A Taxpayer Risk Analysis*. Reason Foundation. <http://reason.org/studies/show/the-xpresswest-high-speed-rail-line> (Accessed August 25, 2012).
- Crompton, J.L. 2006. 'Economic Impact Studies: Instruments for Political Shenanigans?' *Journal of Travel Research* 45: 67–82.
- Danermark, B. et al. 2002. *Explaining Society. Critical Realism in the Social Sciences*. London: Routledge.
- Dempsey, T. 2003. *Delphic Oracle. Its Early History, Influence and Fall*. London: Kessinger Publishing.
- Department for Transport. 2006. *An introduction to variable demand modelling*. London: Integrated Transport Economics and Appraisal (ITEA) Division. Transport Analysis Guidance (TAG).
- DN. 2009. '2000 kroner i støtte - per billett'. *Dagens Næringsliv*. <http://www.dn.no/forsiden/politikkSamfunn/article1684811.ece> (Accessed September 26, 2012).
- Downs, A. 1962. 'The law of peak-hour expressway congestion'. *Traffic Quarterly* 16(3): 393–409.
- DR. 2007. 'Succes for Ringbanen'. *Danmarks Radio*. <http://www.dr.dk/Regioner/Kbh/Nyheder/Hovedstadsomraadet/2007/04/22/085325.htm> (Accessed April 11, 2012).

- DSB. 2004. 'Parkér & Rejs fra Kildedal Station'. <http://www.dsb.dk/om-dsb/presse/pressemeddelelser/parker--rejs-fra-kildedal-station/> (Accessed March 27, 2012).
- . 2008. 'Toget har stor succes'. <http://www.dsb.dk/om-dsb/presse/pressemeddelelser/toget-har-stor-succes/> (Accessed April 11, 2012).
- DtF. 1998. *A new deal for transport: better for everyone*. London: Department for Transport. White paper.
- Duncan, R. 2008. 'Problematic practise in integrated impact assessment: the role of consultants and predictive computer models in burying uncertainty'. *Impact Assessment and Project Appraisal* 26(1): 53–66.
- EC. 2001. *Guidance on EIA*. Edinburgh: European Commission: Environmental Resources Management.
- Flyvbjerg, B. 2005. 'Measuring inaccuracy in travel demand forecasting: Methodological considerations regarding ramp up and sampling'. *Transportation Research Part A: Policy and Practice* 39(6): 522–530.
- . 2007. 'Megaproject Policy and Planning: Problems, Causes, Cures'. Doctoral Dissertation. Aalborg University.
- . 2009. 'Survival of the unfittest'. *Oxford Review of Economic Policy* 25(3): 344–367.
- Flyvbjerg, B., N. Bruzelius, and W. Rothengatter. 2003. *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge: University Press.
- Flyvbjerg, B., M.K.S. Holm, and S.L. Buhl. 2005. 'How (in)accurate are demand forecasts in public works projects?: The case of transportation'. *Journal of the American Planning Association* 71(2): 131–146.
- . 2006. 'Inaccuracy in traffic forecasts'. *Transport Reviews* 26(1): 1–24.
- . 2002. 'Underestimating costs in Public Works Projects: Error or Lie?' *Journal of the American Planning Association* 68(3): 279–296.
- Fouracre, P.R., R.J. Allport, and J.M. Thomson. 1990. *The performance and impact of rail mass transit in developing countries*. UK Transport and Road Research Laboratory.
- FT. 2000. 'Forslag til Lov om anlæg af Ringbanen mellem Hellerup og Ny Ellebjerg'. <https://www.retsinformation.dk/Forms/R0710.aspx?id=90562> (Accessed March 29, 2012).
- Fullerton, B., and S. Openshaw. 1985. 'An evaluation of the Tyneside Metro'. In *International railway economics. Studies in management and efficiency*, eds. K.J. Button and D.E. Pitfield. Aldershot: Gower Publishing Company, p. 177–208.
- Goetzke, F. 2003. *Are Travel Demand Forecasting Models Biased because of Uncorrected Spatial Autocorrelation?* Morgantown: Department of Economics, West Virginia University. Paper presented at the North American Meeting of the Regional Science Association International.

- Goodwin, P. 1996. 'Empirical evidence of induced traffic. A review and synthesis'. *Transportation Research Part A* 33: 655–669.
- Grant-Muller, S.M. et al. 2001. 'Economic appraisal of European transport projects. The state of the art revisited'. *Transport Reviews* 21(2): 237–261.
- Growther, G. 1963. *Traffic in Towns. A study of the long term problems of traffic in urban areas*. London: Reports of the Steering Group and Working Group Appointed by the Minister of Transport.
- Haezendonck, E. 2007. *Transport project evaluation: extending the social cost-benefit approach*. Edward Elgar Publishing.
- Hajer, M.A. 1997. *The Politics of Environmental Discourse: Ecological Modernization and the Policy Process*. Oxford University Press, USA.
- Hall, P. 1980. *Great planning disasters*. London: Weidenfeld & Nicholson.
- Hall, P., and C. Hass-Klau. 1985. *Can Rail Save the City?: Impacts of Rail Rapid Transit and Pedestrianisation on British and German Cities*. Ashgate Publishing Limited.
- Hall, P. 2002. *Urban and Regional Planning Fourth Edition*. 4th ed. Routledge.
- Harman, G.G. 1965. 'The Inference to the Best Explanation'. *The Philosophical Review* 74(1): 88–95.
- Hayashi, Y., and H. Morisugi. 2000. 'International comparison of background concept and methodology of transportation project appraisal'. *Transport Policy* 7(1): 73–88.
- Henman, P. 2002. 'Computer modeling and the politics of greenhouse gas policy in Australia'. *Social Science Computer Review* 20(2): 161–173.
- Hills, P.J. 1996. 'What is induced traffic?'. *Transportation* 23: 5–16.
- Hogarth, R.M., and S. Makridakis. 1981. 'Forecasting and planning. An evaluation'. *Management Science* 27(2): 115–138.
- Holm, M.K.S. 2000. 'Economic appraisal of large scale transport infrastructure investments'. Ph.D. Thesis. Aalborg University.
- IK. 2008. *Danmarks Transportinfrastruktur 2030. Betænkning 1493*. Danish Infrastructure Commission (Infrastrukturkommissionen).
- Jamieson, S. 2004. 'Likert scales: how to (ab)use them'. *Medical Education* 38: 1217–1218.
- Jay, S. et al. 2007. 'Environmental impact assessment. Retrospect and prospect'. *Environmental Impact Assessment Review* 27(4): 287–300.
- Johnston, R.A. 2006. *Review of U.S. and European Regional Modeling Studies of Policies Intended to Reduce Motorized Travel, Fuel Use, and Emissions*. Victoria: Victoria Transport Policy Institute.

- Jones, L.R., and K. Euske. 1991. 'Strategic Misrepresentation in Budgeting'. *Journal of Public Administration Research and Theory* 3: 37–52.
- De Jongh, P. 1998. 'Uncertainty in EIA'. In *Environmental Impact Assessment: Theory and Practice*, ed. Peter Warthern. Routledge.
- Jovicic, G., and C.O. Hansen. 2001. 'The Orestad Traffic Passenger Demand Model'. In *Proceedings of the Annual Traffic Conference at AAU, Aalborg*.
- Jyllands Posten. 2011. 'Borgmestre: Metro under Øresund'. *Jyllands Posten*. <http://jp.dk/indland/trafik/article2639069.ece> (Accessed March 26, 2012).
- Kahneman, D., and A. Tversky. 1979. 'Intuitive predictions. Biases and corrective procedures'. *TIMS Studies in Management Science* 12: 313–327.
- Kain, J.F. 1990. 'Deception in Dallas. Strategic Misrepresentation in Rail Transit Promotion and Evaluation'. *Journal of the American Planning Association* 56(2): 184–196.
- Kharbanda, O.P., and E.A. Stallworthy. 1983. *How to learn from project disasters. True-life stories with a moral for management*. Aldershot: Gower Publishing Company.
- Koslowski. 2011. *Predicts 2012: Technology innovation starts a new automotive era*. San Jose: Gartner.
- Krogh, C. 2011. 'Trafikplanlægning', workshop presentation at Aalborg University.
- Litman, T., and S.B. Colman. 2001. 'Generated Traffic. Implications for Transport Planning'. *Institute of Transportation Engineers Journal* 71(4): 38–47.
- Lovallo, D., and D. Kahneman. 2003. 'Delusions of Success: How Optimism Undermines Executives' Decisions'. *Harvard Business Review* 81(7): 56–63.
- Mackenzie, D. 1993. *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*. The MIT Press.
- Mackett, R.L., and M. Edwards. 1998. 'The impact of new urban public transport systems. Will the expectations be met?'. *Transportation Research A* 32(4): 231–245.
- Mackie, P. 2010. *Cost-Benefit analysis in transport. A UK perspective*. Discussion paper presented at International Transport Forum roundtable, Mexico.
- Mackie, P., S. Jara-Diaz, and A.S. Fowkes. 2001. 'The value of travel time savings in evaluation'. *Transportation Research Part E* 37(2-3): 91–106.
- Mackinder, I.E., and S.E. Evans. 1981. *The predictive accuracy of British transport studies in urban areas*. Crowthorne, Berkshire: Transport and Road Research Laboratory.
- March, James G. 1994. *A primer on decision making*. New York: The Free Press.

- Marfelt, B. 2010. 'Kæmpe regnefal af Cowi sår tvivl om rådgiveres troværdighed'. *Inginøren*. <http://ing.dk/artikel/107707-kaempe-regnefejl-af-cowi-saar-tvivl-om-raadgiveres-trovaerdighed> (Accessed August 8, 2012).
- . 1996. 'Løst overslag fra DSB røg direkte i Folketinget'. *Inginøren*. <http://ing.dk/artikel/14269-loest-overslag-fra-dsb-roeg-direkte-i-folketinget> (Accessed June 4, 2012).
- Marte, G. 2003. 'Slow Vehicle Traffic is a more Attractive Alternative to Fast Vehicle Traffic than Public Transport'. *World Transport Policy & Practice* 9(2): 18–23.
- McNally, M.G. 2007. 'The four step model'. In *Handbook of Transport Modelling*, Elsevier Science.
- Van Melsen, A.G. 1952. *From atomos to atom. The history of the concept atom*. Mineola: Dover.
- Merton, R.K. 1968. *Social Theory and Social Structure*. Free Press.
- Mill, J.S. 1882. *System of Logic*. 8th ed. New York: Harper & Brothers.
- Mitroff, I.I., and A. Silvers. 2010. *Dirty Rotten Strategies. How we trick ourselves into solving the wrong problems precisely*. Stanford, California: Stanford University Press.
- MOA. 1985. *85-rapporten om Storebælt*. Copenhagen: Danish Ministry of Transport (Ministeriet for Offentlige Arbejder).
- Mogridge, M.J.H. 1990. *Travel in towns. Jam yesterday, jam today and jam tomorrow?* London: Macmillan Press.
- MOTOS. 2007. *Transport Modelling: Towards Operational Standards in Europe*. MOTOS project EU. Handbook. http://www.motosproject.eu/?po_id=handbook (Accessed December 18, 2009).
- Murga, M. 2002. 'The 4-step demand model'. <http://ocw.mit.edu/courses/urban-studies-and-planning/11-380j-urban-transportation-planning-fall-2002/lecture-notes/Week5a.pdf>.
- Nakamura, H. 2000. 'The economic evaluation of transport infrastructure: needs for international comparisons'. *Transport Policy* 7(1): 3–6.
- NAO. 1988. *Department of Transport, Scottish Development Department and Welsh Office: Road Planning*. London: National Audit Office.
- Naustdalslid, J., and M. Reitan. 1994. *Kunnskap og styring. Om forskningens rolle i politikk og forvaltning*. Oslo: TANO, NIBR.
- Newman, P., and J. R. Kenworthy. 1989. 'Gasoline consumption and cities: a comparison of US cities with a global survey'. *Journal of the American Planning Association* 55(1): 24–37.
- Nicolaisen, M.S., and P. Næss. 2011. 'Vejenes virkelige værdi'. *Trafik & Veje* 88(10): 44–46.
- Nielsen, O.A., and M. Fosgerau. 2005. 'Overvurderes tidsbenefit af vejprojekter?' In *Proceedings of the Annual Traffic Conference at AAU, Aalborg*.

- Noland, R. B., and L. L. Lem. 2002. 'A review of the evidence for induced travel and changes in transportation and environmental policy in the US and the UK'. *Transportation Research Part D: Transport and Environment* 7(1): 1–26.
- Næss, P. 2006. 'Cost-Benefit Analyses of Transportation Investments: Neither critical nor realistic'. *Journal of Critical Realism* 3(1): 32–60.
- . 2011. 'The Third Limfjord Crossing. A case of pessimism bias and knowledge filtering'. *Transport Reviews* 31(2): 231–249.
- Næss, P., M.J.H. Mogridge, and S.L. Sandberg. 2001. 'Wider roads, more cars'. *Natural Resources Forum* 25: 147–155.
- Næss, P., M.S. Nicolaisen, and A. Strand. 2012. 'Traffic Forecasts ignoring induced demand: a shaky fundament for cost-benefit analyses'. *European Journal of Transport and Infrastructure Research* 12(3).
- Næss, P., and A. Strand. 2012. 'What kinds of traffic forecasts are possible?' *Journal of Critical Realism* 11(3).
- Odeck, J. 2010. 'What Determines Decision-Makers' Preferences for Road Investments? Evidence from the Norwegian Road Sector'. *Transport Reviews* 30(4): 473–494.
- Odgaard, T., C. Kelly, and J. Laird. 2005. *Current practice in project appraisal in Europe. Analysis of country reports*. European Commission EC-DG TREN.
- Olesen, J.H. 2011. 'Magten i modellen', workshop presentation at Aalborg University.
- Ortúzar, J.d.D, and L.G. Willumsen. 2011. *Modelling Transport*. 4th ed. Chichester: John Wiley & Sons.
- Oskamp, S. 1965. 'Overconfidence in case study judgements'. *Journal of Consulting Psychology* 29(3): 261–265.
- Osland, O., and A. Strand. 2010. 'The Politics and Institutions of Project Approval. A Critical-Constructive Comment on the Theory of Strategic Misrepresentation'. *European Journal of Transport and Infrastructure Research*: 77–88.
- Overgaard, K.R. 1966. *Traffic Estimation in Urban Transportation Planning*. Lyngby: Technical University of Denmark.
- Owens, S. 1995. 'From "predict and provide" to "predict and prevent"? Pricing and planning in transport policy'. *Transport Policy* 2(1): 43–49.
- Parkinson, C.N. 1955. 'Parkinson's Law'. *The Economist*.
- Parthasarathi, P., and D. Levinson. 2010. 'Post construction evaluation of traffic forecast accuracy'. *Transport Policy* 12(6): 428–443.
- Pas, E. 1995. 'The urban transportation planning process'. In *The geography of urban transportation*, ed. S. Hanson. New York: Guilford Press, p. 53–77.

- Pickrell, D.H. 1992. 'A desire named streetcar'. *Journal of the American Planning Association* 58(2): 158.
- Pickrell, D.H. 1990. *Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost*. Urban Mass Transportation Administration.
- Pilegaard, A., and B. Johnsen. 1999. 'Ringbanen'. In Aalborg.
- Politiken. 2007. 'Øresundsbro er snart for smal'. *Politiken*.
<http://politiken.dk/udland/ECE220562/oeresundsbro-er-snart-for-smal/> (Accessed March 26, 2012).
- Raiffa, H. 1985. *The Art and Science of Negotiation*. Belknap Press of Harvard University Press.
- Ramjerdi, F. 2011. 'Modellberegninger som selvopfyllende profeti'. *Samferdsel* (8): 20.
- Ramsey, S. 2005. 'Of mice and elephants'. *ITE Journal* September.
- Reed, B. 2000. 'Accidental Truth and Accidental Justification'. *The Philosophical Quarterly* 50(198): 57–67.
- Reichenbach, H. 1949. *The Theory of Probability*. University of California Press.
- RH. 2009. *Udvikling i den kollektive trafik i Region Hovedstaden*. Copenhagen: Danish Capital Region (Region Hovedstaden).
- Rich, J. 2010. *Introduction to transport models. Theory, application and implementation*. Lyngby: Technical University of Denmark.
- Ricoeur, P. 1975. *Freud and Philosophy: An Essay of Interpretation*. New Haven: Yale University.
- Rittel, H.W.J., and M.M. Webber. 1973. 'Dilemmas in a general theory of planning'. *Policy sciences* 4: 155–169.
- RR. 2009. *Beretning til Statsrevisorerne om budgetoverskridelser i statslige bygge- og anlægsprojekter*. Copenhagen: Audit of the State Accounts (Rigsrevisionen).
- Rumsfeld, D.H., and R. Myers. 2002. 'Defense.gov News Transcript: DoD News Briefing - Secretary Rumsfeld and Gen. Myers'.
<http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636> (Accessed January 16, 2012).
- Russell, B. 1967. *A History of Western Philosophy*. Simon & Schuster/Touchstone.
- SACTRA. 1994. *Trunk roads and the generation of traffic*. London: Standing Advisory Committee on Trunk Road Assessment.
- Salling, K.B., and D. Bannister. 2009. 'Assessment of large transport infrastructure projects: the CBA-DK model'. *Transportation Research Part A* 43: 800–813.

- Salling, K.B., and S. Leleur. 2006. 'Assessment of transport infrastructure projects by the use of monte carlo simulation. The CBA-DK model'. *Proceedings of the 2006 Winter Simulation Conference*: 1537–1544.
- Sayer, A. 1992. *Method in social science. A realist perspective*. 2nd ed. London: Routledge.
- Sharot, T., C.W. Korn, and R.J Dolan. 2011. 'How unrealistic optimism is maintained in the face of reality'. *Nature Neuroscience* 14(11): 1475–1481.
- Siemiatycki, M. 2009. 'Academics and Auditors. Comparing Perspectives on Transportation Project Cost Overruns'. *Journal of Planning Education and Research* 29(2): 142–156.
- SINTEF. 2002. *Kompletterende beregninger for analyse av Bybane i Bergen*. Trondheim: SINTEF.
- Sivak, M., and B. Schoettle. 2012. 'Update: Percentage of Young Persons With a Driver's License Continues to Drop'. *Traffic Injury Prevention* 13(4): 341–341.
- Skaburskis, A., and M.B Teitz. 2003. 'Forecasts & outcomes'. *Planning Theory and Practise* 4(4): 429–442.
- Skamris, M.K. 1994. *Large Transport Projects. Forecast vs. Actual Traffic and Cost*. Aalborg: Aalborg University. Master Thesis.
- Slovic, P. 1972. 'Psychological Study of Human Judgment: Implications for Investment Decision Making'. *The Journal of Finance* 27(4): 779–799.
- Soyer, Emre, and R.M. Hogarth. 2012. 'The illusion of predictability: How regression statistics mislead experts'. *International Journal of Forecasting* 28(3): 695–711.
- Tal, G., and G. Cohen-Blankshtein. 2011. 'Understanding the role of the forecast-maker in overestimation forecasts of policy impacts: The case of Travel Demand Management policies'. *Transportation Research Part A* 45: 389–400.
- Taleb, N.N. 2010. *The Black Swan: Second Edition: The Impact of the Highly Improbable: With a new section: 'On Robustness and Fragility'*. 2nd ed. Random House Trade Paperbacks.
- Talvitie, A. 2000. 'Evaluation of road projects and programs in developing countries'. *Transport Policy* 7(1): 61–72.
- TCRP. 2004. *Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects*. Washington: Transportation Research Board.
- Tetraplan. 2011. *Trafikmodel- og effektberegninger for Nordhavnsvej i København*. Internal memo to AAU for contracted consultancy work.
- Thomopoulos, N., S.M. Grant-Muller, and M.R. Tight. 2009. 'Incorporating equality considerations in transport infrastructure evaluation. Current practise and a proposed methodology'. *Evaluation and Program Planning* 32(4): 351–359.
- Thomson, J.M. 1977. *Great cities and their traffic*. London: Gollancz.

- Tichy, G. 2004. 'The over-optimism among experts in assessment and foresight'. *Technological Forecasting & Social Change* 71: 341–363.
- TRACE. 1999. *Elasticity Handbook: Elasticities for Prototypical Contexts*. European Commission, Directorate-General for Transport.
- TRM. 2007. *Notat til Rigsrevisionen: Oversigt over Banedanmarks anlægsprojekter på § 28.63.02 afsluttet i perioden 2003-2007*. Copenhagen: Ministry of Transport (Transportministeriet). Internal Note.
- . 1998. *Trafikal og økonomisk vurdering af anlæg og drift af Ringbanen*. Danish Ministry of Transport (Trafikministeriet). Summary Report.
- TUT. 2011. 'Worldwide winners at the 2011 Light Rail Awards'. *Tramways & Urban Transit* (Special Issue).
- TV. 2011. *Redovisning av samhällsekonomiska efterkalkyler för åtgärder öppnade för trafik 2010, Citytunneln*. Region Syd: Swedish Transport Administration. Accounting Report.
- Tversky, A., and D. Kahneman. 1973. 'Availability: A Heuristic for Judging Frequency and Probability'. *Cognitive Psychology* 5(2): 207–232.
- UNITE. 2011a. *Database of forecast and observed demand and cost for completed transport infrastructure projects*. Aalborg University. Unpublished database from the on-going research project UNITE.
- . 2011b. *Survey data from interviews 2010-2011*. Aalborg University. Unpublished database from the on-going research project UNITE.
- . 2011c. *Survey data from questionnaires 2010-2011*. Aalborg University. Unpublished database from the on-going research project UNITE.
- VD. 2012. *Kommentarer til høringssvar om 3. Limfjordsforbindelse*. Danish Road Directorate (Vejdirektoratet).
- . 1995. 'Notat om anlægsoverslag og trafikprognoser ved større anlæg på hovedlandevejsnettet'.
- . 2003. *Udbygning af Køge Bugt Motorvejen mellem Hundige og Greve Syd*. Danish Road Directorate (Vejdirektoratet).
- VG TAG. 1992. *Transportanalyse Gardermoen*. Verification Group for TA Gardermoen (Verifiseringsgruppen). Verification Report.
- Vickerman, R. 2000. 'Evaluation methodologies for transport projects in the United Kingdom'. *Transport Policy* 7(1): 7–16.
- Wachs, M. 1989. 'When Planners Lie with Numbers'. *Journal of the American Planning Association* 55(4): 476–479.

- Walker, W.E. et al. 2003. 'Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision Support'. *Integrated Assessment* 4(1): 5–17.
- Walmsley, D. A., and M. W. Pickett. 1992. *The cost and patronage of rapid transit systems compared with forecasts*. Research Report 352, Transport Research Laboratory.
- Van Wee, B. 2011. *Transport and Ethics: Ethics and the Evaluation of Transport Policies and Projects*. Edward Elgar Pub.
- Welde, M., and J. Odeck. 2011. 'Do planners get it right? The accuracy of travel demand forecasting in Norway'. *European Journal of Transport and Infrastructure Research* 1(11): 80–95.
- Welin. 2004. 'Station indviet med blå næser'. *BT*. <http://www.b.dk/danmark/station-indviet-med-blaa-naeser> (Accessed March 29, 2012).
- Wessman, J. 2006. 'Ketchupeffekt för Öresundspendlingen'. *Sydsvenskan*. <http://www.sydsvenskan.se/omkretsen/ketchupeffekt-for-oresundspendlingen/> (Accessed March 26, 2012).